

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

**POSGRADO EN CIENCIAS
MATEMÁTICAS**

FACULTAD DE CIENCIAS

**ANÁLISIS DE COMPONENTES PRINCIPALES
COMUNES EN VECTORES DEPENDIENTES BAJO
EL SUPUESTO DE DISTRIBUCIÓN ELÍPTICA
APLICADO A ENCUESTAS LONGITUDINALES
DE DISEÑO MUESTRAL COMPLEJO**

TESIS

QUE PARA OBTENER EL GRADO ACADÉMICO DE

DOCTOR EN CIENCIAS

PRESENTA

HUMBERTO SOTO DE LA ROSA

DIRECTORA DE TESIS: DRA. SILVIA RUIZ-VELASCO ACOSTA

MEXICO, D.F.

NOVIEMBRE, 2007



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Agradecimientos:

A mi amada esposa por todo el apoyo en la realización de este proyecto.

A la doctora Silvia Ruiz-Velasco Acosta, por todo el tiempo invertido, por su gran orientación y por su confianza.

A los doctores Manuel Mendoza, Rubén Hernández, Carlos Cuevas e Ignacio Méndez por sus comentarios y recomendaciones que agregaron un gran valor a esta tesis.

A mis padres y hermana, que me motivaron a nunca conformarme y siempre seguir adelante.

A todos los familiares y amigos de quienes recibí apoyo, tanto material como emocional.

A mi fiel ayudante y compañía en las buenas y en las malas.

A Diosito por sobre todas las cosas.

GRACIAS, MUCHAS GRACIAS!!

Índice:

Motivación y Objetivos

Motivación y Objetivos	1
------------------------------	---

Capítulo I. Antecedentes

I.1 Antecedentes en el Área de Análisis Multivariado	4
I.2 Antecedentes en el Área de Diseño Muestral Complejo	5
I.3 Antecedentes en el Área de Metodología Estadística, Univariada y Multivariada, Contemplando Diseños Muestrales Complejos	6

Capítulo II. Análisis de Componentes Principales Comunes en Vectores Dependientes

II.1 Bases Metodológicas	9
II.2 Modelo del Análisis de Componentes Principales (ACP)	11
II.3 Modelo del Análisis de Componentes Principales Comunes (ACPC)	18
II.4 Modelo del Análisis de Componentes Principales Comunes en Vectores Dependientes (ACPCD)	23

Capítulo III. Distribuciones Elípticas

III.1 Definición de Distribución Elíptica	29
III.2 Teoremas para Distribuciones Elípticas	31
III.3 Ejemplo (Distribución Pearson Tipo VII)	36
III.4 Simulación de Vectores con Distribución Elíptica	41

Capítulo IV. Análisis de Componentes Principales Comunes en Vectores Dependientes Bajo el Supuesto de Distribución Elíptica

IV.1 Estimación Puntual de los Parámetros	53
IV.2 Distribución Asintótica de los Parámetros	63
IV.3 Comparaciones Teóricas con Respecto a los Resultados Bajo el Supuesto de Normalidad	78
IV.4 Comparaciones Empíricas con Respecto a los Resultados Bajo el Supuesto de Normalidad	81

Capítulo V. Análisis de Componentes Principales Comunes en Vectores Dependientes Bajo el Supuesto de Distribución Elíptica en Diseños Muestrales Complejos

V.1 Estimación Puntual de los Parámetros por el Método de Pseudo Máxima Verosimilitud	92
V.2 Estimación Puntual de los Parámetros por el Método de Pseudo Máxima Verosimilitud, en el caso del Análisis de Componentes Principales Comunes en Vectores Dependientes	100
V.3 Inferencia sobre los parámetros en el caso de muestreos complejos	104

Capítulo VI. Aplicación del Análisis de Componentes Principales Comunes en Vectores Dependientes Bajo el Supuesto de Distribución Elíptica en Muestreo Aleatorio Simple

VI.1 Índice de Marginación de México 1990-2000	106
--	-----

Capítulo VII. Aplicación del Análisis de Componentes Principales Comunes en Vectores Dependientes Bajo el Supuesto de Distribución Elíptica en Diseños Muestrales Complejos

VII.1 Encuesta Nacional de los Niveles de Vida de los Hogares, México 2002-2005	116
--	-----

Conclusiones

Comentarios finales y conclusiones	129
--	-----

Bibliografía

Bibliografía	135
--------------------	-----

Apéndice

Programa que implementa el algoritmo FG+	142
--	-----

Motivación y Objetivos

Las técnicas estadísticas de análisis multivariado de datos fueron extensamente estudiadas a lo largo del siglo XX, desarrollándose la teoría de nuevos métodos durante todo el siglo y teniendo un aumento en la frecuencia de su aplicación en la segunda mitad del mismo, debido principalmente al incremento en las facilidades de cálculo computacional y análisis numérico.

La causa principal en el auge del desarrollo de dichas técnicas fué, y aun sigue siendo, la necesidad de estudiar a los individuos de una población, a partir de una muestra representativa, utilizando una visión amplia, que contemple múltiples características de la referida población en un mismo análisis, con el objetivo de resumir la información de las mismas en indicadores sencillos de interpretar. De este modo surgieron las propuestas de Análisis de Regresión Múltiple, de Análisis Multivariado de Datos de Tipo Categórico, y análisis más específicos como las técnicas estadísticas de agrupación (e. g. Análisis de Conglomerados), las técnicas estadísticas de clasificación (e. g. Análisis Discriminante), Análisis Multivariado de Varianza, Análisis de Factores y Análisis de Componentes Principales, por mencionar las más conocidas.

Una de las áreas que han establecido la referida necesidad de utilizar una visión multidimensional, es el área de evaluación de proyectos, debido a la búsqueda de establecer de manera conjunta los indicadores que varían y la forma en que lo hacen, entre dos o más poblaciones de individuos en estudio, o entre una misma población en distintos momentos del tiempo. Hasta ahora la mayoría de los estudios se tienen que limitar a realizar análisis indicador por indicador. Esta es la primera motivación del presente proyecto, teniendo como primer objetivo el presentar los resultados existentes con respecto a una técnica poco conocida, y casi nunca tratada en los libros de texto, denominada Análisis de Componentes Principales Comunes en Vectores Dependientes con la finalidad de plantearla como una alternativa para realizar ese análisis de variación de indicadores de manera conjunta, contemplando la posibilidad de que las poblaciones estudiadas no sean independientes, ya sea por tener características en común, o por formar parte de una encuesta longitudinal. Esto permitirá al área de evaluación de proyectos tener visiones de la evolución de los indicadores de manera conjunta.

Ahora bien, la mayoría de las técnicas estadísticas multivariadas, han basado sus desarrollos en un conjunto de supuestos distribucionales y muestrales, que generalmente consisten en asumir que los datos provienen de una población con distribución Normal Multivariada y que se tiene información captada a partir de una muestra representativa de la población a estudiar, obtenida a partir de muestreo aleatorio simple.

Por una parte, las críticas relacionadas con la dificultad de que, en la práctica, el supuesto distribucional de Normalidad Multivariada se cumpla, han derivado en la necesidad de estudiar distribuciones alternativas o mas generales, y por tanto motivado en parte el desarrollo del presente proyecto, en el que se establece como segundo objetivo, presentar a la familia de distribuciones elípticas, y los resultados estadísticos necesarios para la implementación práctica del Análisis de Componentes Principales Comunes en Vectores Dependientes partiendo del supuesto de distribución elíptica de los datos.

Por otra parte, también en la práctica, la creciente necesidad de estudiar a los individuos de una población a partir de la visión multidimensional, se ha combinado con el aumento en el tamaño y la dispersión de las poblaciones estudiadas, principalmente en lo que a estudios de carácter sociológico y económico se refiere, lo cual ha generado la necesidad de modificar los mecanismos de captación de la información en la búsqueda de optimizar los recursos disponibles. Esto ha implicado que se desarrolle el área de diseño muestral, lo cual se ha venido realizando durante toda la segunda mitad del siglo XX, proponiendose alternativas al muestreo aleatorio simple, denominadas diseños muestrales complejos (estratificados, conglomerados, polietápicos, etc.).

Las consecuencias al utilizar datos obtenidos a partir de diseños muestrales complejos se reflejaron en que los análisis estadísticos existentes no contemplaban las modificaciones derivadas de dichos diseños en sus desarrollos teóricos, por lo cual los resultados obtenidos podrían ser incorrectos. Esto ha provocado la necesidad de retornar a dichos desarrollos teóricos, tanto a niveles univariados como multivariados, para que se incluyan en los mismos las modificaciones que generan las variaciones en los diseños muestrales. Existen avances al respecto, que se han concentrado en los métodos estadísticos de mayor frecuencia de uso, sin embargo aún hay muchos métodos de creciente aplicación que no contemplan las modificaciones derivadas del uso de datos captados a partir de diseños muestrales complejos. Esta es una más de las motivaciones del presente proyecto que establece como su tercer objetivo, el aportar el desarrollo estadístico necesario para que el Análisis de Componentes Principales Comunes en Vectores Dependientes, pueda ser aplicado de manera práctica partiendo del hecho de que se dispone de una muestra obtenida por medio de un diseño muestral complejo.

Para alcanzar los tres objetivos previamente establecidos, el trabajo se estructurará en siete capítulos, siendo el primero una revisión del trabajo existente en relación a cada uno de los tres grandes rubros que se desprenden de los objetivos. El segundo capítulo presenta la técnica estadística denominada Análisis de Componentes Principales Comunes en Vectores

Dependientes, partiendo de sus bases metodológicas y de las técnicas estadísticas de las cuales se derivó. Este capítulo da cumplimiento al primer objetivo del proyecto. El tercer capítulo hace una semblanza de ciertos resultados de carácter general relacionados con la familia de distribuciones elípticas, como lo son su definición, la forma de realizar análisis para muestras aleatorias, y la forma de simular datos con distribución elíptica. Los tres primeros capítulos son totalmente independientes entre sí, lo cual permite estudiar cualquiera de los tres sin necesidad de haber leído los otros.

En el cuarto capítulo se presenta el desarrollo teórico que establece la forma en que el Análisis de Componentes Principales Comunes en Vectores Dependientes, pueda ser aplicado de manera práctica partiendo del supuesto de distribución elíptica de los datos, dando cumplimiento al segundo objetivo del proyecto.

En el quinto capítulo se presenta el desarrollo estadístico mediante el cual el Análisis de Componentes Principales Comunes en Vectores Dependientes, puede ser aplicado de manera práctica partiendo del hecho de que se dispone de una muestra obtenida por medio de un diseño muestral complejo, considerando la posibilidad de asumir distribución elíptica. Este capítulo da cumplimiento al tercer objetivo del proyecto.

Los dos últimos capítulos presentan dos aplicaciones que reflejan el gran potencial de uso de la metodología, la primera cuando se tiene un censo (resultados poblacionales) y la segunda cuando se dispone de una muestra obtenida por medio de un diseño muestral complejo. En el primer caso se trata de un análisis de la evolución de la marginación por localidad en México entre 1990 y 2000, utilizando los resultados censales de ambos años. El segundo caso refiere a un análisis de las variaciones en los niveles de vida de los hogares de México utilizando para ello una encuesta de carácter longitudinal obtenida por medio de un diseño muestral complejo, denominada Encuesta Nacional de los Niveles de Vida de los Hogares, analizando los años 2002 y 2005.

NOTA: Los resultados que son base, o que se utilizan para demostrar los teoremas y proposiciones que conforman la contribución original del presente trabajo, se incluyen ya sea con la referencia a textos donde se pueden encontrar las demostraciones, o con demostraciones completas en aquellos casos en los cuales no se encontró un texto conteniéndolas.

Capítulo I. Antecedentes

A continuación se presentan por separado los antecedentes en materia de desarrollos, tanto teóricos como prácticos, existentes sobre tres temas relacionados con los objetivos del proyecto, mismos que a su vez constituirán las bases del desarrollo del mismo. Los temas son: análisis multivariado, diseño muestral complejo y métodos estadísticos que contemplan el diseño muestral complejo cuando éste se presenta. No se presentan descripciones detalladas de los métodos referidos, pero se incluyen las referencias de los libros y artículos donde se pueden encontrar dichas descripciones a detalle.

I.1 Antecedentes en el Área de Análisis Multivariado

Para el estudio de datos categóricos se han desarrollado técnicas como el Análisis de Correspondencias, técnica de presentación gráfica de la asociación entre las categorías de un conjunto de variables de tipo categórico, propuesto en 1935 por H. O. Hartley quien publicaba bajo su nombre alemán original Hirschfeld (Greenacre, 1984). Otra técnica de estudio de datos categóricos son los Modelos Log-lineales mediante los cuales se intenta reproducir la distribución de una tabla de contingencias. Esta técnica fue producto de los resultados planteados primero de manera separada por G. Udyn Yule y Karl Pearson en 1900, y posteriormente estudiados con mayor claridad por Maurice Bartlett en 1935 (Fienberg, 1979), (Agresti, 1990).

En el caso del estudio de variables de tipo continuo, se tienen múltiples propuestas de análisis como los Modelos de Regresión utilizados tanto para describir relaciones entre un conjunto de variables explicativas y una de respuesta, como para crear modelos de predicción. Se atribuye tanto a Carl Friedrich Gauss como a Adrien Marie Legendre el descubrimiento del método de mínimos cuadrados, pero todo indica que es sir Francis Galton quien introduce el concepto de Regresión, y aunque originalmente utilizó el término de reversión en un artículo enviado a la Royal Institution en 1877, posteriormente utiliza ya el de Regresión en 1885 (Draper, et al., 1998).

Existen desarrollos sobre métodos de agrupación, como el Análisis de Conglomerados Jerárquicos o el de Medias Móviles, así como métodos de clasificación como el uso de la Regresión Logística o el Análisis Discriminante propuesto por R. A. Fisher en 1936, que parten de agrupaciones previamente conformadas, y evalúan dichas agrupaciones para obtener reglas de clasificación de nuevos individuos en los grupos, utilizando para ello la información de sus características (Huberty, 1994), (Kleinbaum, 1992).

Se han desarrollado también métodos de caracterización de estructuras de correlación entre dos conjuntos de variables, como el Análisis de Correlación Canónica propuesto por Harold Hotelling en 1936, que reemplaza los conjuntos de variables originales por dos nuevos conjuntos con un número menor de variables que son altamente correlacionadas por pares (Anderson, 1988).

El desarrollo de los métodos previos y de otros desarrollos prácticos condujeron a la necesidad de crear también métodos de reducción de dimensionalidad como lo son el Análisis de Factores y el Análisis de Componentes Principales. El Análisis de Factores trata de obtener factores latentes mediante un modelo, propuesto por primera ocasión por Charles Spearman en 1904, mientras que el Análisis de Componentes Principales construye combinaciones lineales ortogonales a partir de un conjunto de variables. La propuesta original de esta técnica fué hecha en su enfoque geométrico por Karl Pearson en 1901 y en su enfoque algebraico por Harold Hotelling en 1933 (Muirhead, 1982), (Anderson, 1988).

I.2 Antecedentes en el Área de Diseño Muestral Complejo

En los años recientes, se ha generado una creciente necesidad de utilizar diseños muestrales complejos para la captación de la información sobre los individuos de una población, entendiendo por complejo a aquél diseño que no se basa en un muestreo aleatorio simple. Dentro de los diseños de muestreo complejo se encuentran los muestreos estratificados, los muestreos por conglomerados, y muestreos polietápicos, cuyas principales características en común son que las probabilidades de selección de cada individuo no son iguales, y que se introducen dependencias probabilísticas entre los individuos. Son justamente esas diferencias las causantes que los análisis estadísticos tanto univariados como multivariados, que no contemplan el diseño muestral, puedan generar resultados erróneos.

La teoría estadística que involucra la construcción de diseños muestrales complejos ha sido extensamente estudiada (Cochran, 1963; Kish, 1965), y en la actualidad se aplica en la captación de información de muy diversas áreas. En dicha teoría se define como efecto de diseño al cociente de las varianzas de los estimadores, obtenidas al contemplar el diseño muestral complejo, y las obtenidas cuando no se contempla el diseño muestral complejo .

Es importante señalar que para que un análisis estadístico pueda contemplar el hecho de que los datos provienen de una encuesta de diseño complejo, es necesario conocer tanto las probabilidades de selección de cada individuo muestreado, como su pertenencia a estratos o

conglomerados en cada etapa de selección. En la práctica, los recíprocos de las probabilidades de selección se incluyen en las bases de datos con el nombre de ponderadores o factores de expansión. También se incluyen generalmente las pertenencias de los individuos a estratos y/o conglomerados, sin embargo éstos elementos se utilizan con poca frecuencia, aplicándose a dichas bases de datos análisis estadísticos que no contemplan el diseño muestral complejo, ya sea por desconocimiento o por falta de disponibilidad de métodos que lo contemplen.

Para concluir con los antecedentes relativos al diseño muestral se hace mención de que, como parte de los desarrollos teóricos que involucran la construcción de los diseños muestrales, se ha estudiado ampliamente la manera de corregir las muestras que por motivos de inexacta declaración o de altas tasas de no respuesta de manera no aleatoria, ya no sean representativas de la población que se desea analizar. Para ello se utilizan métodos de calibración de ponderadores (Deville, et al. 1996), y esto puede llevar a que aún una encuesta sin diseño muestral complejo deba ser analizada por los métodos estadísticos que contemplen el diseño muestral complejo, que en este caso no será más que el efecto por calibración.

I.3 Antecedentes en el Área de Metodología Estadística, Univariada y Multivariada, Contemplando Diseños Muestrales Complejos

Varios métodos estadísticos tanto descriptivos como inferenciales, univariados y multivariados contemplan ya el diseño muestral complejo. El desarrollo teórico existente para esos métodos puede ser utilizado para incorporarlo a otros métodos estadísticos que no lo contemplan. En ello radica la importancia de presentar los antecedentes de este tema.

Desde que se desarrolló la teoría estadística que involucra la construcción de diseños muestrales complejos, lo cual ocurrió poco después de la mitad del siglo XX, se han venido generando también formas de considerar el efecto de diseño en los análisis estadísticos. Esto se complica a medida que se pasa de análisis estadísticos inferenciales simples como pueden ser estimaciones puntuales, intervalos de confianza o pruebas de hipótesis, a análisis que involucran modelos en su diseño. En cualquier caso, los análisis que involucran modelos requieren también de estimaciones, intervalos o pruebas de hipótesis para generar las conclusiones estadísticas correspondientes.

Se tiene ya un amplio nivel de estudio en lo que concierne a la consideración del diseño muestral complejo en los métodos estadísticos que no contemplan modelos, concentrándose en estadísticos descriptivos e inferencias en intervalos de confianza y pruebas de hipótesis (Kish, et al., 1974).

En lo que respecta a modelos específicos, se han realizado los desarrollos teóricos correspondientes a modelos de regresión (Nathan, et al., 1980; Graubard, et al., 1994; Binder, 1992, etc.). Para datos categóricos también existen desarrollos (Rao, et al. 1981). Para otros análisis multivariados también se tienen referencias que contemplan el diseño muestral complejo (Koch, et al., 1972; Koch, et al., 1975; Bebbington, et al., 1977; Pervaiz, et al., 1992; Skinner, et al., 2005). En el caso específico del análisis de Componentes Principales las referencias existentes se concentran en la correcta estimación de los componentes y de sus pesos (Skinner, et al. 1986; Skinner, et al. 1989).

I.3.1 Herramientas estadísticas para considerar el diseño muestral complejo en los estimadores

En materia de estimadores, los métodos que se han considerado para incorporar el diseño muestral complejo son principalmente:

i) Mínimos cuadrados ponderados por la probabilidad de selección: estimación que no se basa en un supuesto distribucional, utilizando la metodología de mínimos cuadrados pero considerando que los individuos tienen pesos distintos (Skinner, et al. 1989).

ii) Pseudo máxima verosimilitud, basado en encontrar el estimador que maximiza la verosimilitud pero solo de entre un conjunto de posibles estimadores que no constituyen el universo total de los posibles estimadores. Esto puede hacerse asumiendo una distribución nominal aproximada a la distribución real de los datos, o por medio de la estimación de funciones de los parámetros que son más sencillas de estimar que los parámetros mismos (Skinner, et al. 1989).

iii) Ajuste de Pearson a los estimadores máximo verosímiles, consistente en incorporar un factor de ajuste a los estimadores obtenidos sin considerar el diseño complejo, de tal modo que corrijan el sesgo generado por dicho diseño (Skinner, et al. 1989).

iv) Estimación Bayesiana, partiendo de la estructura conceptual y teórica planteada desde la perspectiva del análisis bayesiano de datos, como por ejemplo, por medio de la aplicación del concepto de función de utilidad (Lee, 1989; Robert, 2001).

I.3.2 Herramientas estadísticas para considerar el diseño muestral complejo en la estimación de las varianzas o errores estándar de los estimadores

En lo que concierne a la estimación de las varianzas o errores estándar de los estimadores considerando el efecto de diseño, los métodos más difundidos y utilizados son:

i) Estimación paramétrica (Modelo bajo el supuesto de una distribución parametrizable conocida). En este caso se trata de llegar a la distribución de los estimadores, ya sea por medio de su obtención analítica, o por medio del uso de resultados que permitan aproximarla de manera asintótica. Una vez con la referida distribución se pueden analizar las correspondientes varianzas (Serfling, 1980).

ii) Estimación no paramétrica que comprende entre otros a los métodos de Estimaciones Lineales, Replicación Simple, Replicación balanceada, Jackknife y Bootstrap (Skinner, et al. 1989; Wolter, 1985).

Capítulo II. Análisis de Componentes Principales Comunes en Vectores Dependientes

En este capítulo se presenta la descripción de la técnica del Análisis de Componentes Principales Comunes y la forma de utilizarla para analizar encuestas longitudinales. Para ello se presentan primero una serie de resultados necesarios para la comprensión de la teoría que sustenta la técnica, posteriormente se presenta la técnica de Análisis de Componentes Principales, después se muestra la manera de generalizar dicha técnica para aplicarla a un estudio por grupos por medio del Análisis de Componentes Principales Comunes y finalmente se presentan su extensión para utilizarla en grupos dependientes, en particular en el caso del análisis de encuestas longitudinales.

II.1 Bases metodológicas

Definición II.1. Sea $A \in \mathbb{R}^{p \times p}$ una matriz cuadrada, $\underline{x} = (x_1, \dots, x_p) \in C^p$ es un vector propio de A si $\underline{x} \neq \underline{0}$ y $A\underline{x} = \lambda\underline{x}$ para alguna $\lambda \in C$ denominada valor propio.

La ecuación $A\underline{x} = \lambda\underline{x}$ puede escribirse equivalentemente como: $(A - \lambda I_p)\underline{x} = \underline{0}$, y si \underline{x} es solución no trivial, entonces debe ocurrir que $c(\lambda) = \det(A - \lambda I_p) = 0$. La función c es un polinomio de grado p en λ denominado polinomio característico y tiene exactamente p raíces en C . Puede ocurrir que algunas raíces tengan multiplicidad mayor a 1. Los vectores propios \underline{x} asociados a los valores propios λ_j de multiplicidad mayor a 1, no están definidos de manera única por lo que comúnmente se establecen ciertas restricciones para obtenerlos.

Si hay multiplicidad de valores propios mayor a 1, los vectores propios se escogen de entre los vectores del espacio asociado al valor propio (que es de dimensión igual a la multiplicidad) y que sean ortogonales.

Lema II.1. Los valores propios de una matriz real simétrica son todos reales.

Demostración: se encuentra en [Magnus, et. al., 1999, pag. 14].

Lema II.2. Los vectores propios asociados a dos valores propios distintos de una matriz simétrica son ortogonales.

Demostración: se encuentra en [Magnus, et. al., 1999, pag. 16].

El siguiente teorema, denominado de Diagonalización y Descomposición Espectral de Matrices Simétricas, garantiza la existencia de los valores y vectores propios cuando la matriz es simétrica.

Teorema II.1. Sea $A \in \mathbb{R}^{p \times p}$ una matriz simétrica ($A' = A$). Entonces existe una matriz ortogonal β ($\beta'\beta = \beta\beta' = I_p$) y una matriz diagonal Λ tales que:

$$\beta' A \beta = \Lambda \quad (II.1)$$

Si los valores propios son todos distintos entonces β está definida de manera única salvo por multiplicación de columnas por -1 y permutación de columnas.

Demostración: se encuentra en [Magnus, et. al., 1999, pag. 16].

Es claro que al premultiplicar la igualdad $\beta' A \beta = \Lambda$ por β se obtiene $A\beta = \beta\Lambda$, de donde se aprecia que las columnas de β , es decir $(\underline{\beta}_1, \dots, \underline{\beta}_p)$, son los vectores propios de A normalizados ($\beta'_j \underline{\beta}_j = 1, j = 1, \dots, p$), y los elementos de la diagonal Λ son los valores propios asociados.

Si la igualdad $A\beta = \beta\Lambda$ es postmultiplicada por β' , se obtiene $A = \beta\Lambda\beta' = \sum_{j=1}^p \lambda_j \underline{\beta}_j \underline{\beta}'_j$. Dicha forma de escribir la igualdad se conoce como descomposición espectral.

Definición II.2. Una matriz $A \in \mathbb{R}^{p \times p}$ simétrica es positiva definida si para todo vector $\underline{x} \in \mathbb{R}^{p \times p}$ se tiene que $\underline{x}' A \underline{x} > 0$.

Lema II.3. Si todos los valores propios de una matriz simétrica A son positivos, entonces la matriz es positiva definida ($\underline{x}' A \underline{x} > 0$).

Demostración: se encuentra en [Magnus, et. al., 1999, pag. 15].

Lema II.4. Si todos los valores propios de una matriz simétrica A son mayores o iguales a cero, entonces la matriz es positiva semidefinida ($\underline{x}' A \underline{x} \geq 0$).

Demostración: se encuentra en [Magnus, et. al., 1999, pag. 15].

Es importante notar además que si A es una matriz simétrica positiva definida entonces se tiene que, dado que $A = \sum_{j=1}^p \lambda_j \underline{\beta}_j \underline{\beta}'_j$ con $\lambda_j > 0$, entonces A^α para $\alpha \in \mathbb{R}$ es:

$$A^\alpha = \sum_{j=1}^p \lambda_j^\alpha \underline{\beta}_j \underline{\beta}'_j = \beta \Lambda^\alpha \beta'$$

donde $\Lambda^\alpha = \text{diag}(\lambda_1^\alpha, \dots, \lambda_p^\alpha)$.

Otro teorema de utilidad para plantear la posibilidad de combinar la diagonalización en matrices distintas, buscando elementos en común es el Teorema de Diagonalización Simultánea de dos Matrices Simétricas que se enuncia a continuación.

Teorema II.2. Sean $A \in \mathbb{R}^{p \times p}$ y $W \in \mathbb{R}^{p \times p}$ matrices simétricas y suponga que A es positiva definida, entonces existe una matriz no singular G y una diagonal Λ tales que:

$$G'AG = I_p \tag{II.2}$$

y

$$G'WG = \Lambda \tag{II.3}$$

Demostración: se encuentra en [Magnus, et. al., 1999, pag. 22].

Conjuntando las ecuaciones (II.2) y (II.3) del **Teorema II.2**, se puede escribir una tercera ecuación de la siguiente forma: $G'WG = G'AG\Lambda$ misma que premultiplicada por G'^{-1} es $WG = A\Lambda$. Por ser A una matriz simétrica y positiva definida, existe A^{-1} , así es que se puede premultiplicar la igualdad anterior por dicha matriz, y entonces se obtiene $A^{-1}WG = G\Lambda$. Escribiendo $H = G'^{-1}$ se tiene que $A = HH' = \sum_{j=1}^p h_j h'_j$ y $W = H\Lambda H' = \sum_{j=1}^p \lambda_j h_j h'_j$.

II.2 Modelo del Análisis de Componentes Principales (ACP)

En 1901 Karl Pearson desde un enfoque geométrico, y posteriormente en 1933 Harold Hotelling desde un enfoque algebraico, plantearon el método conocido como Análisis de Componentes Principales, que actualmente es una de las herramientas Análisis Multivariado con mayores aplicaciones, pues permite reducir la dimensionalidad de los datos así como identificar grupos de variables cuya información se refiere a un mismo concepto.

El Análisis de Componentes Principales transforma un conjunto de variables correlacionadas en otro conjunto de variables no correlacionadas que son combinaciones lineales de las primeras, que se denominan Componentes Principales, y que están ordenadas en el sentido de la variabilidad que representan con respecto a las originales.

II.2.1 Cálculo teórico de las Componentes Principales

Cuando se dispone de la información de toda la población de individuos, el Análisis de Componentes Principales para los vectores p -variados \underline{x} conteniendo la información de p variables para cada individuo de una población, consiste en obtener primero una combinación lineal $u_1 = \underline{a}'_1 \underline{x}$ con $\underline{a}_1 \in \mathbb{R}^p$ tal que la varianza de u_1 sea lo mas grande posible, con una restricción adicional sobre \underline{a}_1 puesto que sin dicha restricción u_1 puede ser arbitrariamente grande. La restricción más comúnmente utilizada es: $\underline{a}'_1 \underline{a}_1 = 1$.

Una vez definida la primera combinación lineal, que se denomina Primera Componente Principal, se procede a encontrar una segunda combinación lineal $u_2 = \underline{a}'_2 \underline{x}$ tal que la varianza de u_2 sea lo mas grande posible sujeto a la restricción de que $\underline{a}'_2 \underline{a}_2 = 1$, pero condicionalmente a la existencia de la Primera Componente Principal, por lo cual se incorpora como restricción adicional $\underline{a}'_1 \underline{a}_2 = 0$, implicando la ortogonalidad entre la Primera y Segunda Componentes Principales. El procedimiento continúa de manera sucesiva hasta obtener $\underline{u}' = (u_1, \dots, u_p)$ componentes principales con varianza decreciente y ortogonales entre sí.

Teorema II.3. Sea Ψ la matriz de varianzas y covarianzas respectiva de un conjunto de vectores \underline{x} p -variados conteniendo la información de p variables medidas a una población de individuos, encontrar los vectores \underline{a}_i que maximizan sucesivamente la varianza de $u_i = \underline{a}'_i \underline{x}$ sujeto a las restricciones $\underline{a}'_i \underline{a}_i = 1$ y $\underline{a}'_j \underline{a}_i = 0$, es equivalente a encontrar las columnas de la matriz β ortogonal tal que para una matriz Λ diagonal se satisface la ecuación:

$$\Psi = \beta \Lambda \beta' \quad (II.4)$$

En este caso los valores $u_i = \underline{a}'_i \underline{x}$ que son llamados Componentes Principales, son no correlacionados entre sí y que pueden ordenarse en términos de la varianza explicada por medio de los valores diagonales de Λ , debido a que $Var(\underline{u}) = \Lambda$, para $\underline{u}' = (u_1, \dots, u_p)$

Demostración: se encuentra en [Flury, 1988, pag. 8].

Hasta este punto se ha presentado la forma de encontrar las Componentes Principales cuando se conoce la matriz de varianzas y covarianzas poblacional. En la práctica es frecuente

el estudio de muestras de individuos, para los cuales sólo se puede obtener una estimación de la matriz de varianzas y covarianzas $\hat{\Psi}$. En estos casos es necesario recurrir a métodos estadísticos de estimación, lo cual conlleva la necesidad de realizar inferencias sobre los resultados para evaluar el margen de error. Adicionalmente al error de estimación se deberá sumar el error por cálculo computacional, debido a que en la práctica en la mayoría de los casos se requerirá el uso de computadoras para obtener las Componentes Principales. Estos dos temas se abordan a continuación.

II.2.2 Métodos Estadísticos de Estimación de las Componentes Principales

Cuando sólo se dispone de una muestra de tamaño n de la población que se desea analizar, es necesario realizar estimaciones de los componentes principales, desde un punto de vista estadístico. Este rubro ha sido extensamente estudiado para el caso en que se supone que la muestra ha sido obtenida mediante un Muestreo Aleatorio Simple (MAS) y cuando se supone también que el vector de variables (ahora aleatorio) tiene una distribución Normal Multivariada, es decir, $\underline{x} \sim N(\underline{\mu}, \Psi)$, donde se plantea a $\hat{\Psi} = S$ conociéndose que la distribución de $(n-1)S$ es Wishart ($(n-1)S \sim W(\Psi, n-1)$) donde $S = \frac{1}{n-1} \sum_{i=1}^n (\underline{x}_i - \bar{\underline{x}})(\underline{x}_i - \bar{\underline{x}})'$. En este caso, $\bar{\underline{x}}$ tiene distribución Normal Multivariada, $\bar{\underline{x}}$ y S son las estadísticas suficientes para $\underline{\mu}$ y Σ , y son independientes. (Anderson, 1988).

Teorema II.4. Los estimadores máximo verosímiles $\hat{\beta}$ y $\hat{\Lambda}$ de las matrices β y Λ correspondientes al modelo de Análisis de Componentes Principales, bajo el supuesto de distribución Normal Multivariada, corresponden a los vectores y valores propios de la matriz $\frac{n-1}{n}S$ respectivamente, es decir, satisfacen la ecuación:

$$\frac{n-1}{n}S = \hat{\beta}\hat{\Lambda}\hat{\beta}' \quad (II.5)$$

Demostración: se encuentra en [Flury, 1988, pag. 14].

Para realizar inferencia estadística sobre los estimadores de los valores y vectores propios de la matriz de varianzas y covarianzas muestral S , se puede recurrir a la inferencia asintótica, y de esa forma se pueden generar intervalos de confianza y pruebas de hipótesis para los valores y vectores propios de la matriz Ψ , cuando se cuenta con muestras suficientemente grandes.

Lema II.5. La distribución asintótica del m -ésimo elemento de la diagonal de la matriz $\widehat{\Lambda}$ denotado como $\widehat{\lambda}_m$ está dada por:

$$\widehat{\lambda}_m \sim N\left(\lambda_m, \frac{2\lambda_m^2}{n}\right) \quad (11.6)$$

Se puede usar $s(\widehat{\lambda}_m) = \sqrt{\frac{2}{n}}\widehat{\lambda}_m$ como estimación del error estandar.

Demostración: se encuentra en [Flury, 1988, pag. 26].

Lema II.6. La distribución asintótica del vector (m) -ésimo de la matriz $\widehat{\beta}$ denotado como $\widehat{\underline{\beta}}_m$, bajo el supuesto de que $\lambda_i \neq \lambda_j$, está dada por:

$$\widehat{\underline{\beta}}_m \sim N(\underline{\beta}_m, v_m) \quad (11.7)$$

donde

$$v_m = \frac{\lambda_m}{n} \sum_{i=1, i \neq m}^p \frac{\lambda_i}{(\lambda_i - \lambda_m)^2} \underline{\beta}_i \underline{\beta}_i' \quad (11.8)$$

con $\widehat{\underline{\beta}}_i$ y $\widehat{\underline{\beta}}_j$ independientes ($i \neq j$). Se puede usar como estimador del error estándar:

$$s(\widehat{\underline{\beta}}_m) = \left(\frac{\widehat{\lambda}_m}{n} \sum_{i=1, i \neq m}^p \frac{\widehat{\lambda}_i}{(\widehat{\lambda}_i - \widehat{\lambda}_m)^2} \widehat{\underline{\beta}}_i \widehat{\underline{\beta}}_i' \right)^{\frac{1}{2}} \quad (11.9)$$

Demostración: se encuentra en [Flury, 1988, pag. 30].

II.2.3 Métodos computacionales para el cálculo de las Componentes Principales

El teorema de descomposición espectral de matrices simétricas garantiza la existencia de las matrices β y Λ con las características antes descritas dado que Ψ es una matriz simétrica y positiva definida. Debido que se han encontrado muchas aplicaciones en diversos campos de estudio, para el problema de álgebra matricial consistente en encontrar las matrices β ortogonal y Λ diagonal que satisfacen la ecuación $\Psi = \beta \Lambda \beta'$ para alguna matriz Ψ simétrica y positiva definida, existen varios algoritmos computacionales eficientes que permiten encontrar en la práctica aproximaciones numéricas de las matrices β y Λ . A continuación se presentarán algunas de las propuestas, una breve descripción de ellas y un análisis comparativo entre las mismas.

El algoritmo QR consiste en diagonalizar una matriz simétrica premultiplicándola y postmultiplicándola por una matriz ortogonal y su correspondiente transpuesta respectivamente. Para encontrar la matriz ortogonal, el algoritmo se subdivide en dos etapas que operan, ambas, de manera iterativa. La primera etapa logra que por medio de la construcción de matrices de Householder se obtenga una matriz tridiagonal y simétrica, conocida como matriz de Hessenberg. Es importante señalar que una matriz de Householder es una matriz ortogonal que permite eliminar un elemento de otra matriz al premultiplicarla y postmultiplicarla por la primera. Una vez obtenida la matriz de Hessenberg, el algoritmo pasa a su segunda etapa, que consiste en eliminar los elementos que están fuera de la diagonal, reduciéndolos a cero o a una cantidad cercana a cero. Para ello se pueden utilizar dos métodos, el de cambio explícito por medio de lo que se denomina cambios de Wilkinson, o la versión de cambio implícito. Este algoritmo requiere aproximadamente $9p^3$ operaciones (Golub y Van Loan, 1989).

Otra alternativa para diagonalizar una matriz simétrica premultiplicándola y postmultiplicándola por una matriz ortogonal y su correspondiente transpuesta respectivamente, es el método de Jacobi que se basa en el uso de las matrices denominadas rotaciones de Givens, que son distintas a las matrices de Householder pero tienen exactamente el mismo efecto de eliminación de elementos de una matriz al pre y postmultiplicar por ellas, además de ser también matrices ortogonales. El algoritmo de Jacobi se basa en generar dichas rotaciones de Givens de manera iterativa hasta lograr que los elementos fuera de la diagonal de la matriz estudiada sean todos lo más cercano a cero. Si bien al eliminar un elemento puede ser que otro elemento previamente eliminado vuelva a aparecer, se puede demostrar que en conjunto la suma de los elementos fuera de la diagonal siempre se reduce con cada iteración. Se puede probar también que si el elemento fuera de la diagonal que se elige para ser eliminado en una iteración es el mayor, entonces se obtiene la mayor disminución en la suma del resto de los elementos, sin embargo computacionalmente, encontrar el elemento mayor implica un costo que es preferible evitar, para ello se ha propuesto un algoritmo de manera cíclica, lo cual implica que se elige un ordenamiento preestablecido para seleccionar a los elementos a eliminar (Golub y Van Loan, 1989).

En general el algoritmo de Jacobi es inferior en cuanto a rapidez de convergencia con respecto al algoritmo QR, sin embargo su programación es mucho más simple y para p (dimensión de la matriz) pequeñas puede ser competitivo, además cuando la matriz analizada es casi diagonal, el algoritmo de Jacobi puede utilizar esa información mientras que el QR no. Adicionalmente, desde la aparición del cómputo en paralelo, el algoritmo de Jacobi puede tener ventajas computacionales con respecto al QR con pequeñas modificaciones en el algoritmo, por ello es que ambos algoritmos pueden ser utilizados esperando rapidez de cómputo similar en dimensiones pequeñas (Golub y Van Loan, 1989).

Otros algoritmos disponibles pero menos utilizados porque son apropiados solo cuando se desea obtener algunos valores y/o vectores propios son: el cociente de iteración de Rayleigh, el método de bisección y el método de iteración con aceleración de Ritz. Mención aparte merece otro método denominado de Lanczos, de difícil aplicación práctica debido a los errores de redondeo que llegan a producir pérdida de ortogonalidad en los vectores propios calculados (Golub y Van Loan, 1989).

Está demostrado que los métodos antes señalados convergen, sin embargo no se debe descuidar nunca la posibilidad de que la naturaleza de los datos afecte su eficiencia (análisis de sensibilidad). Asimismo se debe tener presente que los resultados obtenidos son de carácter numérico, lo cual implica que acarrearán errores de redondeo que en el peor de los casos podrían conducir, por el fenómeno de la cancelación catastrófica, a resultados totalmente distintos a los reales, por ello se debe tratar de realizar análisis de error. (Golub y Van Loan, 1989).

II.2.4 Problemas comunes al realizar un Análisis de Componentes Principales

Cuando se realiza un Análisis de Componentes Principales puede surgir una serie de problemas o dificultades que es necesario considerar y en su caso corregir. Entre dichos problemas se encuentra la violación al supuesto de normalidad multivariada que sostiene las estimaciones estadísticas basadas en el concepto de verosimilitud. Una forma de corregir este problema es buscando una distribución distinta o más general que la normal que pueda modelar a los datos y que sea manejable, una opción puede ser una distribución elíptica. Otra forma de corregir el problema es por medio del uso de estadística no paramétrica. Esta dificultad no es importante cuando se desea realizar solamente un análisis descriptivo, para lo cual se pueden calcular las componentes principales sin necesidad de utilizar verosimilitudes, tomando en cuenta el hecho de que se parte de la matriz S y de que se busca una transformación lineal que maximice la varianza de los datos.

Otro punto que tiene que ser considerado con atención es la violación del supuesto de que todos los valores propios son diferentes en lo que se conoce como problema de esfericidad. Existen pruebas para validar el supuesto, pero en caso de que las pruebas indiquen su violación se debe poner especial cuidado tanto en la interpretación como en el cálculo numérico de los valores y vectores propios, pues algunos métodos numéricos de estimación tienen problemas al enfrentarse a una matriz con valores propios muy similares o iguales (ver **Lema II.6**).

Un tercer problema que puede aparecer al realizar el Análisis de Componentes Principales es el que los datos tengan ciertas observaciones influyentes (outliers) que pueden llevar a que los resultados obtenidos no sean correctos, tanto en el caso de que dichos datos se hayan debido a errores, por ejemplo de transcripción, como en el caso de que los datos influyentes impliquen la falta de normalidad.

Un cuarto problema es el que se conoce como atrición que consiste en que las variables tengan valores faltantes. El problema es complicado debido a que los valores faltantes entre distintas variables no necesariamente corresponden a los mismos individuos, por lo cual el cálculo de las covarianzas no se puede realizar de manera directa. Para corregir el problema se pueden imputar a los valores faltantes las medias de las variables respectivas, o generar un procedimiento para obtener un dato mejor aproximado por medio del uso de análisis de regresión, o en un caso extremo, eliminar el registro, lo cual pudiera reducir considerablemente el tamaño de muestra. También se pueden calcular por separado las entradas de la matriz de varianzas y covarianzas, eliminando sólo los valores faltantes de las variables requeridas en el cálculo de cada entrada. Algunas de las alternativas presentadas pueden conducir a problemas posteriores al tratar de obtener las componentes principales debido a que la matriz de varianzas y covarianzas puede no ser positiva definida. Adicionalmente, persiste el problema de calcular los componentes de los individuos con datos faltantes.

En un Análisis de Componentes Principales puede presentarse la necesidad de seleccionar sólo a las componentes principales más informativas, lo cual puede hacerse evaluando su contribución. Una manera muy utilizada para ello es por medio de una medida construida como el cociente de la suma de los mayores valores propios en orden descendente entre la suma total de los valores propios. Otra necesidad que puede aparecer es la selección de variables debido a que no necesariamente todas las variables introducidas al análisis son útiles. Esta necesidad puede evaluarse midiendo las correlaciones entre las variables originales y las componentes principales, buscando detectar aquellas variables que o bien no estén relacionadas con ninguna componente o bien no estén relacionadas con las componentes principales de mayor contribución. Alternativamente se pueden seleccionar las variables útiles por el valor de las β 's, recordando que si hay supuestos distribucionales, se puede hacer inferencia sobre los valores de las mismas, o incluso de las contribuciones.

Finalmente se debe mencionar que en ocasiones se utiliza la matriz de correlación en vez de la matriz de covarianzas, ya sea por cuestión de disponibilidad o por buscar ventajas como el tratar de eliminar efectos diferenciales por las escalas utilizadas, por ejemplo. Cuando esto ocurre no se pueden realizar inferencias a menos que éstas se realicen sobre las variables centradas y escaladas, y en caso contrario solo se podrá realizar un análisis descriptivo.

II.3 Modelo del Análisis de Componentes Principales Comunes (ACPC)

En ciertas aplicaciones en donde se requiere analizar a más de un conjunto de individuos, aplicar el Análisis de Componentes Principales por separado a cada conjunto no garantiza la comparabilidad entre las Componentes Principales de los conjuntos, porque en el caso en que las varianzas entre conjuntos sean distintas, se obtienen combinaciones lineales diferentes para cada conjunto. Esto motivó a Bernhard Flury (1988) a proponer el modelo denominado Análisis de Componentes Principales Comunes con la finalidad de encontrar las Componentes Principales que son comunes a dos o más conjuntos independientes de individuos, buscando comparar entre conjuntos, los valores de las componentes para definir a partir de dicha comparación los atributos generales que los diferencian.

II.3.1 Cálculo teórico de las Componentes Principales Comunes

El modelo del Análisis de Componentes Principales Comunes consiste en extender el Análisis de Componentes Principales. Para tal fin, establece que, si es posible encontrar una matriz ortogonal β tal que las matrices de varianzas y covarianzas de cada grupo (Ψ_i) se pueden expresar como:

$$\Psi_i = \beta \Lambda_i \beta', \quad i = 1, \dots, k \quad (II.10)$$

con Λ_i matrices diagonales, entonces las columnas de β se pueden utilizar del mismo modo que en el Análisis de Componentes Principales como los coeficientes de las combinaciones lineales de las variables originales que generan a su vez un conjunto de Componentes Principales Comunes a todos los conjuntos, de modo que son comparables y además pueden ordenarse por su contribución a la variabilidad original (Flury, 1988).

Para calcular la matriz β y las matrices $\Lambda_1, \dots, \Lambda_k$ se puede recurrir a procedimientos denominados de diagonalización simultánea de matrices. El sustento matemático para el caso en que $k = 2$ se basa en el teorema de diagonalización simultánea de matrices que se presentó en la **Sección II.1** y que establece la existencia de una matriz no singular que ahora llamaremos β^* y una diagonal Λ^* tales que para dos matrices simétricas y positivas definidas ahora renombradas como Ψ_1 y Ψ_2 , cumplen con $\beta^{*'} \Psi_1 \beta^* = I_p$ y $\beta^{*'} \Psi_2 \beta^* = \Lambda^*$. En la referida sección se encontró a partir del **Teorema II.2** el resultado que a continuación se reescribe adecuado a la notación que ahora se utiliza: $\Psi_1^{-1} \Psi_2 \beta^* = \beta^* \Lambda^*$.

La relación entre las matrices β y β^* así como la existente entre Λ y Λ^* no se puede establecer de manera directa. Sin embargo si se supone que la hipótesis planteada por el Análisis

de Componentes Principales Comunes se cumple, esto es, si $\Psi_1 = \beta \Lambda_1 \beta'$ y $\Psi_2 = \beta \Lambda_2 \beta'$. Estas dos igualdades se pueden sintetizar en una sola de la siguiente manera: $\Psi_1^{-1} \Psi_2 \beta = \beta \Lambda_1^{-1} \Lambda_2$, donde la matriz $\Lambda_1^{-1} \Lambda_2$ es diagonal. Por el resultado descrito en el párrafo anterior, es posible que existan $\beta^* = \beta$ y $\Lambda^* = \Lambda_1^{-1} \Lambda_2$, sin embargo este resultado no garantiza la existencia. Flury demuestra un resultado que establece la condición para garantizar dicha existencia, y es que $\Psi_1^{-1} \Psi_2$ sea simétrica (Flury, 1988).

Para encontrar las matrices Λ_i y β Flury partió del concepto de diagonalización simultánea, para lo cual propuso una medida de desviación de la diagonalización simultánea.

Definición II.3. Sean F_1, \dots, F_k matrices simétricas y positivas definidas, y n_1, \dots, n_k pesos asociados a las matrices, una medida de desviación de la diagonalización simultánea es:

$$\Phi(F_1, \dots, F_k; n_1, \dots, n_k) := \prod_{i=1}^k \left[\log \left(\frac{|diag(F_i)|}{|F_i|} \right) \right]^{n_i} \quad (II.11)$$

Donde $diag(F_i)$ es la matriz con los mismos elementos de F_i en la diagonal, y ceros fuera de ella. Esta definición se sustenta en que, por medio de la desigualdad de Hadamard, se puede demostrar que $\Phi(F_1, \dots, F_k; n_1, \dots, n_k) \geq 1$, cumpliéndose la igualdad cuando existe la diagonalización simultánea. A partir dicha medida y de métodos de optimización, Flury encontró la forma de determinar matrices Λ_i y β .

Teorema II.5. Los valores de Λ_i y β que minimizan la medida de desviación de la diagonalización simultánea $\Phi(\Lambda_1, \dots, \Lambda_k; n_1, \dots, n_k)$ para $\Lambda_i = \beta' \Psi_i \beta$ con Ψ_i matrices simétricas y positivas definidas, satisfacen el siguiente sistema de ecuaciones, sujetos a la restricción de ortogonalidad de β , y bajo el supuesto de que $\lambda_{im} \neq \lambda_{ij}$:

$$\lambda_{im} = \underline{\beta}'_m \Psi_i \underline{\beta}_m, \quad i = 1, \dots, k, \quad m = 1, \dots, p \quad (II.12)$$

$$\underline{\beta}'_m \left(\sum_{i=1}^k n_i \frac{\lambda_{im} - \lambda_{ij}}{\lambda_{im} \lambda_{ij}} \Psi_i \right) \underline{\beta}_j = 0, \quad m, j = 1, \dots, p, \quad m \neq j \quad (II.13)$$

Donde λ_{im} es el m -ésimo elemento de la diagonal de Λ_i y $\underline{\beta}_m$ es la m -ésima columna de β .

Demostración: se encuentra en [Flury, 1988, pag. 68].

II.3.2 Métodos Estadísticos de Estimación de las Componentes Principales Comunes

La frecuencia con que el análisis se debe realizar a partir de muestras, motivó a Flury a derivar una forma de estimar estadísticamente las matrices Λ_i y β , para lo cual supuso normalidad multivariada y obtuvo los estimadores máximo verosímiles.

Teorema II.6. Los estimadores máximo verosímiles de Λ_i y β bajo el supuesto de normalidad multivariada de los datos, para el modelo $\Psi_i = \beta \Lambda_i \beta'$ con Ψ_i matrices de varianzas y covarianzas, satisfacen el siguiente sistema de ecuaciones, sujetos a la restricción de ortogonalidad impuesta sobre β y suponiendo que $\hat{\lambda}_{im} \neq \hat{\lambda}_{ij}$:

$$\hat{\lambda}_{im} = \hat{\beta}'_m S_i \hat{\beta}_m, \quad i = 1, \dots, k, \quad m = 1, \dots, p \quad (II.14)$$

$$\hat{\beta}'_m \left(\sum_{i=1}^k n_i \frac{\hat{\lambda}_{im} - \hat{\lambda}_{ij}}{\hat{\lambda}_{im} \hat{\lambda}_{ij}} S_i \right) \hat{\beta}_j = 0, \quad m, j = 1, \dots, p, \quad m \neq j \quad (II.15)$$

Donde $\hat{\lambda}_{im}$ es el m -ésimo elemento de la diagonal de $\hat{\Lambda}_i$ y $\hat{\beta}_m$ es la m -ésima columna de $\hat{\beta}$ y S_i es la matriz de varianzas y covarianzas muestral.

Demostración: se encuentra en [Flury, 1988, pag. 68].

Vale la pena mencionar que existen también algunas propuestas alternativas de estimación como la que usa la técnica de mínimos cuadrados (Beaghen, 1997).

No se puede asegurar en general la existencia de las Componentes Principales Comunes, esto significa que no se puede tener certeza de encontrar matrices β ortogonal y Λ_i diagonales tales que se cumpla que $\Psi_i = \beta \Lambda_i \beta'$ para $i = 1, \dots, k$. Sin embargo siempre se pueden encontrar matrices β ortogonal y F_i no necesariamente diagonales tales que se cumpla que $\Psi_i = \beta F_i \beta'$ para $i = 1, \dots, k$. El hecho de que se cumpla el supuesto planteado por el Análisis de las Componentes Principales Comunes implica que, para alguna matriz β , las matrices F_i que se encuentren de la expresión $\Psi_i = \beta F_i \beta'$, son todas diagonales.

Flury plantea una prueba estadística, basada en el método de razón de verosimilitudes para determinar la existencia de las Componentes Principales Comunes. La estadística propuesta se construye a través del planteamiento de la hipótesis nula $H_0 : F_i \text{ diagonales}$ y la hipótesis alternativa $H_a : \text{alguna } F_i \text{ no diagonal}$ en el modelo $\Psi_i = \beta F_i \beta'$ con β ortogonal construyendo:

$$X_{ACPC}^2 = -2 \log \frac{L(\widehat{\beta}_1 \widehat{\beta}_1', \dots, \widehat{\beta}_k \widehat{\beta}_k')}{L(\widehat{\Psi}_1, \dots, \widehat{\Psi}_k)} \quad (11.16)$$

De acuerdo con la teoría de pruebas de razón de verosimilitudes, esta estadística tiene distribución asintótica χ^2 con $(k-1) \frac{p(p-1)}{2}$ grados de libertad (Flury, 1988).

Cuando se asume el supuesto distribucional de Normalidad Multivariada de los datos, el estimador máximo verosímil sin restricciones de Ψ_i es $\widehat{\Psi}_i = \frac{n-1}{n} S_i$. De acuerdo a lo establecido por el **Teorema II.6**, se determina que el estimador máximo verosímil de F_i es $\widehat{F}_i = \text{diag}(\widehat{\Lambda}_i)$. Esto implica que la estadística propuesta por Flury se obtiene como:

$$X_{ACPC}^2 = \sum_{i=1}^k n_i \left[\log \left(\frac{|\text{diag}(\widehat{\Lambda}_i)|}{|\widehat{\Lambda}_i|} \right) \right] \quad (11.17)$$

Una demostración más detallada se encuentra en [Flury, 1988, pag. 70]. En esa misma referencia se incluyen pruebas para modelos parciales, que pueden ser útiles en el caso de no validarse la existencia de las Componentes Principales Comunes.

A partir del mismo supuesto de normalidad multivariada, Flury obtuvo la distribución asintótica de los estimadores $\widehat{\Lambda}_i$ y $\widehat{\beta}$, que se presentan en los siguientes lemas.

Lema II.7. La distribución asintótica del m -ésimo elemento de la diagonal de la matriz $\widehat{\Lambda}_i$ denotado como $\widehat{\lambda}_{im}$ está dada por:

$$\widehat{\lambda}_{im} \sim N \left(\lambda_{im}, \frac{2\lambda_{im}^2}{n_i} \right) \quad (11.18)$$

Se puede usar $s(\widehat{\lambda}_{im}) = \sqrt{\frac{2}{n_i}} \widehat{\lambda}_{im}$ como estimación del error estándar.

Demostración: se encuentra en [Flury, 1988, pag. 76].

Lema II.8: La distribución asintótica de la entrada (m, l) de la matriz $\hat{\beta}$ denotado como $\hat{\beta}_{ml}$ está dada por:

$$\hat{\beta}_{ml} \sim N\left(\beta_{ml}, \frac{v_{ml}}{n}\right) \quad (11.19)$$

donde bajo el supuesto de que $\lambda_{ij} \neq \lambda_{il}$:

$$\frac{v_{ml}}{n} = \sum_{j=1, j \neq l}^p \left[\sum_{i=1}^k n_i \frac{(\lambda_{ij} - \lambda_{il})^2}{\lambda_{ij}\lambda_{il}} \right]^{-1} \beta_{mj}^2 \quad (11.20)$$

Se puede usar como estimador del error estándar a:

$$s(\hat{\beta}_{ml}) = \left(\sum_{j=1, j \neq l}^p \left[\sum_{i=1}^k n_i \frac{(\hat{\lambda}_{ij} - \hat{\lambda}_{il})^2}{\hat{\lambda}_{ij}\hat{\lambda}_{il}} \right]^{-1} \hat{\beta}_{mj}^2 \right)^{\frac{1}{2}} \quad (11.21)$$

Demostración: se encuentra en [Flury, 1988, pag. 79].

II.3.3 Métodos computacionales para el cálculo de las Componentes Principales Comunes

Para encontrar las soluciones del sistema definido por los **Teoremas II.5 y II.6**, Flury y Constantine desarrollaron un algoritmo que denominaron algoritmo FG y que consiste en realidad en dos algoritmos anidados, el exterior que va generando de manera cíclica submatrices de dimensión 2×2 a partir de las matrices de dimensión original, y el interior resuelve el sistema planteado para el conjunto de submatrices de 2×2 previamente generadas utilizando para ello rotaciones de Givens (Flury y Constantine, 1984). Se puede apreciar claramente que el algoritmo hace uso de las ideas del Algoritmo de Jacobi con recorrimiento cíclico para encontrar valores y vectores propios.

Se debe destacar que para todo algoritmo numérico propuesto hay que evaluar su estabilidad, probar su convergencia, así como realizar un análisis de sensibilidad (magnitud del cambio en los resultados a cambios pequeños en los valores originales de las matrices). Flury demuestra que el Algoritmo FG es numéricamente estable y que converge (Flury, 1988), sin embargo no realiza análisis de sensibilidad.

II.3.4 Problemas comunes al realizar un Análisis de Componentes Principales Comunes

Existen muy pocas adiciones a los problemas que se presentan al realizar un Análisis de Componentes Principales, el principal problema que se puede presentar es la invalidez de la hipótesis planteada por el Análisis de Componentes Principales Comunes, es decir, que en realidad no existan dichas Componentes Principales Comunes. Para este problema existen como alternativas los modelos parciales de Componentes Principales Comunes, que suponen que solo algunas de las Componentes son comunes entre los grupos, o los modelos de proporcionalidad que suponen que entre las matrices de varianzas y covarianzas existe una proporcionalidad (Flury, 1988). Existen algunos trabajos relacionados con la solución al problema de detección de outliers (Gu, et. al., 2001; Boente, et. al. 2002).

II.4 Modelo del Análisis de Componentes Principales Comunes en Vectores Dependientes (ACPCD)

Existe una limitante al Análisis de Componentes Principales Comunes que se manifiesta cuando los conjuntos de individuos que se quiere analizar no son independientes entre sí, como puede ocurrir en estudios de carácter longitudinal. Para poder considerar esos casos, Beat E. Neuenschwander (1991) desarrolló, a partir del modelo de Flury, la generalización del Análisis de Componentes Principales Comunes que llamó Análisis de Componentes Principales Comunes en Vectores Dependientes, planteando la existencia de una matriz β ortogonal tal que las matrices de varianzas y covarianzas Ψ_{ij} se pueden representar como:

$$\Psi_{ij} = \beta \Lambda_{ij} \beta', \quad i, j = 1, \dots, k \quad (II.22)$$

con Λ_{ij} matrices diagonales (ortogonalidad de las componentes entre conjuntos), donde Ψ_{ij} representa a la matriz de varianzas y covarianzas entre el conjunto i y el conjunto j . (Neuenschwander, 1991).

II.4.1 Cálculo teórico de las Componentes Principales Comunes en Vectores Dependientes

El modelo propuesto previamente plantea la necesidad de diagonalizar un conjunto de matrices simultáneamente, por lo cual la forma de encontrar las matrices β y Λ_{ij} es análoga a la

utilizada en el caso del análisis de componentes principales comunes que cumplen el supuesto de independencia entre grupos. Para encontrar los valores de β y de Λ_{ij} , Neuenschwander partió de la medida de diagonalización simultánea definida por Flury.

Teorema II.7. Los valores de Λ_{ij} y β que para los cuales se minimiza la medida

$$\Phi(\Lambda_{11}, \dots, \Lambda_{1k}, \Lambda_{21}, \dots, \Lambda_{2k}, \dots, \Lambda_{k1}, \dots, \Lambda_{kk}; n_{11}, \dots, n_{1k}, n_{21}, \dots, n_{2k}, \dots, n_{k1}, \dots, n_{kk})$$

en donde las matrices Λ_{ij} quedan determinadas por $\Lambda_{ij} = \beta' \Psi_{ij} \beta$, satisfacen el siguiente sistema de ecuaciones, sujetos a la restricción de ortogonalidad de β :

$$\lambda_{ij,m} = \underline{\beta}'_m \Psi_{ij} \underline{\beta}_m, \quad i, j = 1, \dots, k, \quad m = 1, \dots, p \quad (II.23)$$

$$\underline{\beta}'_m \left(\sum_{i=1}^k \sum_{j=1}^k (\lambda^{ij,l} - \lambda^{ij,m}) \Psi_{ij} \right) \underline{\beta}_l = 0, \quad m, l = 1, \dots, p, \quad (m \neq l) \quad (II.24)$$

Donde $\underline{\beta}_m$ es la m -ésima columna de β y $\lambda^{ij,h}$ son los elementos de Λ^{ij} correspondientes a $\Lambda^{-1} = [\Lambda^{ij}]_{i,j=1,\dots,k}$, definida como:

$$\Lambda^{-1} = [\Lambda^{ij}]_{i,j=1,\dots,k} = \begin{bmatrix} \Lambda^{11} & \dots & \Lambda^{1k} \\ \vdots & \ddots & \vdots \\ \Lambda^{k1} & \dots & \Lambda^{kk} \end{bmatrix}$$

Demostración: se encuentra en [Neuenschwander y Flury, 1994, pag. 184].

II.4.2 Métodos Estadísticos de Estimación de las Componentes Principales Comunes en Vectores Dependientes

Neuenschwander y Flury se basaron nuevamente en el supuesto de normalidad multivariada de los datos y para $n_{ij} = n$ utilizaron el método de máxima verosimilitud para estimar los valores de Λ_{ij} y β , obteniendo por medio del uso de algunos resultados de calculo diferencial matricial y de álgebra matricial, que los estimadores referidos satisfacen un sistema de ecuaciones análogo al encontrado previamente por Flury para el Análisis de Componentes Principales Comunes (Neuenschwander y Flury, 2000).

Teorema II.8. Los estimadores máximo verosímiles de Λ_{ij} y β bajo el supuesto de normalidad multivariada de los datos, para el modelo $\Psi_{ij} = \beta \Lambda_{ij} \beta'$ con Ψ_{ij} matrices de varianzas y covarianzas, satisfacen el siguiente sistema de ecuaciones, bajo el supuesto de que $\hat{\lambda}^{ij,l} \neq \hat{\lambda}^{ij,m}$:

$$\hat{\lambda}_{ij,m} = \hat{\beta}'_m \frac{W_{ij}}{n} \hat{\beta}_m, \quad i, j = 1, \dots, k, \quad m = 1, \dots, p \quad (11.25)$$

$$\hat{\beta}'_m \left(\sum_{i=1}^k \sum_{j=1}^k (\hat{\lambda}^{ij,l} - \hat{\lambda}^{ij,m}) \frac{W_{ij}}{n} \right) \hat{\beta}_l = 0, \quad m, l = 1, \dots, p, \quad (m \neq l) \quad (11.26)$$

sujeto a la restricción de ortogonalidad de $\hat{\beta}$, donde $\hat{\lambda}^{ij,h}$ son los elementos de $\hat{\Lambda}^{ij}$ correspondientes a $\hat{\Lambda}^{-1} = [\hat{\Lambda}^{ij}]_{i,j=1,\dots,k}$, $\hat{\beta}_m$ es la m -ésima columna de $\hat{\beta}$ y $\frac{W_{ij}}{n} = \frac{n-1}{n} S_{ij}$.

Demostración: se encuentra en [Neuenschwander y Flury, 2000, pag. 168].

En este caso, la prueba de razón de verosimilitudes para determinar la existencia de las Componentes Principales Comunes en Vectores Dependientes plantea que, para las matrices $\Psi_{11}, \Psi_{12}, \dots, \Psi_{kk}$, bajo el supuesto de que se cumple el modelo $\Psi_{ij} = \beta F_{ij} \beta'$ con β ortogonal, se prueben las hipótesis $H_0 : F_{ij}$ diagonales contra $H_a : \text{alguna } F_{ij} \text{ no diagonal}$, donde la estadística de prueba resulta ser:

$$\chi^2_{ACPC} = -2 \log \frac{L(\hat{\beta} \hat{F}_{11} \hat{\beta}', \hat{\beta} \hat{F}_{12} \hat{\beta}', \dots, \hat{\beta} \hat{F}_{kk} \hat{\beta}')}{L(\hat{\Psi}_{11}, \hat{\Psi}_{12}, \dots, \hat{\Psi}_{kk})} \quad (11.27)$$

De acuerdo con la teoría de pruebas de razón de verosimilitudes, esta estadística tiene distribución asintótica χ^2 con $(k^2 - 1)p(p - 1)$ grados de libertad.

Bajo el supuesto de Normalidad Multivariada de los datos, la estadística anterior se obtiene como:

$$\chi^2_{ACPCD} = n \left[\log \left(\frac{|pdiag(\hat{\Lambda})|}{|\hat{\Lambda}|} \right) \right] \quad (11.28)$$

siendo $\hat{\Lambda} = [\hat{\Lambda}_{ij}]_{i,j=1,\dots,k}$ y $pdiag(\hat{\Lambda}) = [diag(\hat{\Lambda}_{ij})]_{i,j=1,\dots,k}$, donde bajo el supuesto de que se cumple el modelo de Componentes Principales Comunes en Vectores Dependientes, el estimador máximo verosímil de Λ_{ij} es $\hat{\Lambda}_{ij}$, siendo $\hat{F}_{ij} = diag(\hat{\Lambda}_{ij})$, y donde $\hat{\beta}$ es el correspondiente estimador máximo verosímil de β .

La demostración de un caso más general se encuentra en el **Capítulo 4**.

Neuenschwander y Flury también encontraron las distribuciones asintóticas de los estimadores $\widehat{\Lambda}_{ij}$ y $\widehat{\beta}_m$. En el caso particular de $\widehat{\Lambda}_{ij}$ utilizaron las matrices de conmutación $I_{l,m}$ (Henderson y Searle, 1979) para transformar la matriz $\Lambda = [\Lambda_{ij}]_{i,j=1,\dots,k}$ por medio de la ecuación $\Lambda^* = I_{p,k}\Lambda'_{p,k}$, en una matriz diagonal por bloques de la forma:

$$\Lambda^* = \begin{bmatrix} \Lambda_1^* & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \Lambda_2^* & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \Lambda_p^* \end{bmatrix}$$

de manera que es equivalente encontrar la distribución asintótica de los elementos de $\widehat{\Lambda}^*$ y la de $\widehat{\Lambda}$.

La razón de introducir la matriz $\widehat{\Lambda}^*$ es que las submatrices $\widehat{\Lambda}_h^*$ tienen elementos funcionalmente independientes, mientras que la matriz original $\widehat{\Lambda}$ tiene algunos elementos que son funcionalmente dependientes del resto. Esto se debe a que la suma de los valores propios de una matriz es igual a la traza de la matriz.

Teorema II.9. La distribución asintótica del vector $vech(\widehat{\Lambda}_h^*)$ está dada por:

$$vech(\widehat{\Lambda}_h^*) \sim N_{\frac{k(k+1)}{2}} \left(vech(\Lambda_h^*), \frac{2H_k(\Lambda_h^* \otimes \Lambda_h^*)H'_k}{n} \right) \quad (II.29)$$

donde \otimes representa al producto de Kronecker, $vech$ es el operador que transforma una matriz simétrica en un vector al poner las columnas resultantes de eliminar los elementos superiores a la diagonal de la matriz una debajo de otra, y la matriz H_k satisface la ecuación $vech(A) = H_k vec(A)$ para cualquier matriz simétrica A de dimensión k , siendo vec el operador que transforma una matriz en un vector al poner sus columnas una debajo de otra.

Demostración: se encuentra en [Neuenschwander y Flury, 2000, pag. 172].

Así como la matriz $\widehat{\Lambda}$ tiene algunos elementos que son funcionalmente dependientes del resto, del mismo modo $\widehat{\beta}$ tiene únicamente $\frac{\rho(\rho-1)}{2}$ elementos funcionalmente independientes, por lo cual se define al vector $\widehat{\beta}^*$ como aquel conformado por una elección de $\frac{\rho(\rho-1)}{2}$ elementos de la matriz $\widehat{\beta}$. La causa de que existan elementos funcionalmente dependientes son las restricciones de ortogonalidad impuestas a las columnas de la matriz.

Teorema II.10. La distribución asintótica de la matriz $\hat{\underline{\beta}}^*$ está dada por:

$$\hat{\underline{\beta}}^* \sim N_{\frac{p(p-1)}{2}} \left(\underline{\beta}^*, \frac{\underline{\Sigma}}{n} \right) \quad (II.30)$$

donde

$$\underline{\Sigma}^{-1} = \left[\frac{\partial}{\partial \underline{\beta}^*} \text{vec}(\underline{\beta})' \right] (I_p \otimes \underline{\beta}) \left[\sum_{i=1}^k \sum_{j=1}^k (\Lambda^{ij} \otimes \Lambda_{ij}) - kI_{p^2} \right] (I_p \otimes \underline{\beta}') \left[\frac{\partial}{\partial \underline{\beta}^*} \text{vec}(\underline{\beta})' \right]' \quad (II.31)$$

Demostración: se encuentra en [Neuenschwander y Flury, 2000, pag. 173].

En el teorema anterior se utilizan la notaciones:

$$\frac{\partial}{\partial \underline{y}} \underline{x}' = \begin{bmatrix} \frac{d}{dy_1} x_1 & \cdots & \frac{d}{dy_1} x_n \\ \vdots & \ddots & \vdots \\ \frac{d}{dy_m} x_1 & \cdots & \frac{d}{dy_m} x_n \end{bmatrix}$$

y

$$\left[\frac{\partial}{\partial \underline{y}} \underline{x}' \right]' = \frac{\partial}{\partial \underline{y}'} \underline{x}$$

La derivada parcial del vector $\text{vec}(\underline{\beta})'$ con respecto al vector $\underline{\beta}^*$ se puede obtener resolviendo el sistema de ecuaciones que se genera al derivar implícitamente, sobre cada uno de los elementos de $\underline{\beta}^*$, las ecuaciones generadas por las restricciones de ortogonalidad impuestas sobre las columnas de la matriz $\underline{\beta}$. En este sentido, para obtener la varianza asintótica de todos los elementos de la matriz $\underline{\beta}$ se propone escoger a tres diferentes vectores $\underline{\beta}^*$, uno conteniendo todos los elementos inferiores a la diagonal de la matriz, un segundo conteniendo a los elementos superiores a la diagonal y un tercero conteniendo a los elementos de la diagonal. Esto funciona bien para $p > 2$. Para $p = 2$ basta con escoger dos vectores $\underline{\beta}^*$ que en realidad tendrán un solo elemento de la matriz $\underline{\beta}$ cada vez. Para una descripción con mayor detalle se puede recurrir al Apéndice de [Neuenschwander y Flury, 2000, pag. 176].

II.4.3 Métodos computacionales para el cálculo de las Componentes Principales Comunes en Vectores Dependientes

Neuenschwander y Flury también generaron dos algoritmos para obtener de manera numérica, las componentes principales comunes en vectores dependientes (Neuenschwander

y Flury, 1994). El primer algoritmo denominado como algoritmo FG+ ortogonal que consiste análogamente al algoritmo FG en dos subalgoritmos anidados, el F+ y el G+, ambos iterativos. Es de hecho una generalización directa del algoritmo FG y por tanto del algoritmo con recorrimiento cíclico de Jacobi, razón por la cual se omite la descripción a detalle. Neuenschwander y Flury presentan una prueba de convergencia.

El segundo algoritmo propuesto por Neuenschwander y Flury se denomina algoritmo FG+ no ortogonal que se diferencia del algoritmo FG+ ortogonal porque, justo como su nombre lo indica, no requiere que la matriz β a encontrar sea ortogonal. Este algoritmo tiene ligeras variaciones con respecto al algoritmo FG+ ortogonal, y a pesar de que en el artículo referido no se presenta prueba de convergencia, se menciona que empíricamente ha resultado efectivo.

En el Apéndice se presenta el programa en Fortran que implementa el Algoritmo FG+ ortogonal.

II.4.4 Problemas comunes al realizar un Análisis de Componentes Principales Comunes en Vectores Dependientes

Es claro que al ser el Análisis de Componentes Principales Comunes en Vectores Dependientes una extensión más general del Análisis de Componentes Principales Comunes y a su vez también del Análisis de Componentes Principales, hereda de ellos sus principales problemas de una manera extendida puesto que la dimensión en que se trabaja es mayor. A dichos problemas se debe añadir de una manera muy importante el costo computacional que puede tener el obtener los Componentes buscados.

Un problema adicional es la inexistencia de los Componentes Principales Comunes, en cuyo caso se puede recurrir a la teoría desarrollada por Flury, referente a los modelos parciales, que involucran la posibilidad de plantear que solo una parte de las componentes sean comunes (las más importantes) (Flury, 1988).

Capítulo III. Distribuciones Elípticas

En este capítulo se define a la familia de Distribuciones Elípticas y se presentan algunos resultados importantes relacionados a esta familia de distribuciones. Se presenta un ejemplo y se muestra la forma de obtener vectores simulados con Distribución Elíptica.

III.1 Definición de Distribución Elíptica

Definición III.1. Un vector aleatorio $\underline{x} \in \mathbb{R}^p$ tiene distribución elíptica, que se denota como $\underline{x} \sim EC_p(\underline{\mu}, \Sigma, g)$, cuando para una función dada $g(t)$, la forma funcional de la distribución del vector aleatorio es:

$$\begin{aligned} f(\underline{x}) &= |\Sigma|^{-\frac{1}{2}} g[(\underline{x} - \underline{\mu})' \Sigma^{-1} (\underline{x} - \underline{\mu})] \\ &= |\Sigma|^{-\frac{1}{2}} g(V) \end{aligned} \tag{III.1}$$

donde V es una forma cuadrática que involucra al vector aleatorio \underline{x} , por lo cual es una variable aleatoria univariada.

La distribución Normal Multivariada es una distribución elíptica, debido a que su función de densidad corresponde a la expresión (III.1) cuando $g(t) = (2\pi)^{-\frac{p}{2}} e^{-\frac{1}{2}t}$. Se sabe que en este caso la variable aleatoria V tiene distribución χ^2 con p grados de libertad, y se puede apreciar que existe una relación entre la forma de la función $g(t)$ y la forma funcional de la distribución χ^2 . Establecer la relación entre ambas formas funcionales permitirá dar un sentido más claro a la elección de la función $g(t)$ en términos de la función de distribución de V . Para encontrar esa relación de manera general, se requiere usar el concepto de distribución esférica.

Definición III.2. Un vector aleatorio $\underline{y} \in \mathbb{R}^p$ tiene distribución esférica si $\underline{y} \sim EC_p(\mathbf{0}, I_p, g)$, es decir, si para una función dada $g(t)$, la forma funcional de la distribución del vector aleatorio es $f(\underline{y}) = g(\underline{y}'\underline{y})$.

El nombre de distribución esférica se debe a que en este tipo de distribuciones, las curvas de nivel siempre corresponden a esferas, mientras que por su parte, en el caso de las distribuciones elípticas las curvas de nivel siempre corresponden a elipsoides.

Para un vector con distribución esférica, se tiene que $V = R^2 = \underline{y}'\underline{y} = \|\underline{y}\|^2$ y se puede demostrar en ese caso (Fang y Anderson, 1990a) que la variable aleatoria V tiene la siguiente forma distribucional:

$$f(v) = \frac{\pi^{\frac{p}{2}}}{\Gamma(\frac{p}{2})} v^{(\frac{p}{2})-1} g(v) \quad (III.2)$$

Entonces, para generar una distribución esférica, se puede partir de proponer una distribución para V con la forma de la ecuación (III.2) a partir de la cual se identifique la función $g(t)$. Esto implica que en el caso de las distribuciones esféricas, determinar la función $g(t)$ se puede realizar definiendo la distribución de V (o de R).

Para poder aplicar este resultado al planteamiento de generación de distribuciones elípticas, se propone que si la variable aleatoria V está definida por $V = R^2 = \underline{y}'\underline{y} = \|\underline{y}\|^2$ donde \underline{y} es un vector aleatorio con distribución esférica, entonces para $AA' = \Sigma$ se tiene que la distribución de $\underline{x} = \underline{\mu} + A\underline{y}$ es elíptica, esto es que $\underline{x} \sim EC_p(\underline{\mu}, \Sigma, g)$.

El resultado anterior establece la forma en que se pueden proponer las distribuciones elípticas. Además, se puede demostrar que la distribución de \underline{y} es la misma que la de RU , siendo U un vector uniforme sobre la esfera unitaria de dimensión p . Esto implica que la distribución de \underline{x} es la misma que la de $\underline{\mu} + RAU$ (Fang y Anderson, 1990a).

Para generar propuestas concretas de distribuciones elípticas, se puede partir del planteamiento de distribuciones para V alternativas a la χ^2 con p grados de libertad que genera a la Normal Multivariada. Una primera opción es el planteamiento de una distribución Gamma para V , que se propone a partir de la idea de que la distribución χ^2 es un caso particular de la Gamma. Se puede demostrar que en este caso, la función $g(t)$ toma la siguiente forma:

$$g(t) = \alpha \pi^{-\frac{p}{2}} \beta^{\frac{2\gamma+p-2}{2\alpha}} \frac{\Gamma(\frac{p}{2})}{\Gamma(\frac{2\gamma+p-2}{2\alpha})} t^{\gamma-1} e^{-\beta t^\alpha}, \quad \alpha, \beta > 0 \quad y \quad 2\gamma + p > 2$$

Las distribuciones elípticas así generadas se conocen con el nombre de distribuciones tipo Kotz. La distribución Exponencial Multivariada es un caso particular que se obtiene cuando $\alpha > 0$, $\beta = \frac{1}{2}$ y $\gamma = 1$, y a su vez contiene a la Distribución Laplace o Doble exponencial cuando $\alpha = \frac{1}{2}$, $\beta = \frac{1}{2}$ y $\gamma = 1$ y a la Distribución Normal cuando $\alpha = 1$, $\beta = \frac{1}{2}$ y $\gamma = 1$. Cabe destacar que la función $g(t)$ es decreciente cuando $\gamma = 1$, característica que tomará especial interés en resultados posteriores.

Otra opción que se ha propuesto para la distribución de V es la Beta, para la cual se puede demostrar que la función $g(t)$ toma la siguiente forma:

$$g(t) = \pi^{-\frac{p}{2}} \frac{\Gamma\left(\frac{p+2\gamma+2}{2}\right)}{\Gamma(\gamma+1)} (1-t)^\gamma, \quad 0 \leq t \leq 1 \quad y \quad \gamma > -1$$

Las distribuciones elípticas así generadas se conocen con el nombre de distribuciones Pearson Tipo II. En este caso la función $g(t)$ es decreciente cuando $\gamma > 0$.

Una tercer opción propuesta para la distribución de V es la Beta tipo II. En esta alternativa se puede demostrar que la función $g(t)$ toma la siguiente forma:

$$g(t) = \alpha \frac{\Gamma(m)}{\Gamma\left(m - \frac{p}{2\alpha}\right)} \frac{\Gamma\left(\frac{p}{2}\right)}{\Gamma\left(\frac{p}{2\alpha}\right)} \left(\frac{\alpha}{2\alpha-1\gamma}\right)^{\frac{p}{2\alpha}} \left(1 + \frac{\alpha t}{2\alpha-1\gamma}\right)^{-m}, \quad \alpha, \gamma > 0 \quad y \quad 2m > \frac{p}{2\alpha}$$

Las distribuciones elípticas así generadas se conocen con el nombre de distribuciones Pearson Tipo VII Generalizada. Casos particulares de esta distribución son la correspondiente t de Student Multivariada cuando $m = \frac{\gamma+p}{2}$, y como caso límite la Exponencial Multivariada cuando $m = \frac{\gamma+p}{2}$ y $\gamma \rightarrow \infty$ que como ya se señaló, contiene a la Normal y a la Doble Exponencial Multivariadas. En este caso, la función $g(t)$ siempre es decreciente. Más adelante se ejemplifica a detalle la distribución Pearson Tipo VII correspondiente al caso en que $\alpha = 1$.

Existen otras alternativas para generar distribuciones elípticas, destacándose de estos resultados que la familia de distribuciones elípticas es más general que la distribución Normal Multivariada, y que el uso de distintas distribuciones para la variable aleatoria V generarán distribuciones alternativas a la Normal Multivariada. Además, se puede demostrar que, en caso de existir, la esperanza de un vector con distribución elíptica corresponde a $E(\underline{x}) = \underline{\mu}$, y la varianza a $Var(\underline{x}) = \Psi = \frac{E(V)}{p} \Sigma$, con lo cual se resalta la importancia de partir de la definición de la distribución de V .

III.2 Teoremas para Distribuciones Elípticas

A continuación se presenta una serie de teoremas que permiten que se pueda realizar inferencia estadística sobre los parámetros de las Distribuciones Elípticas prácticamente con el mismo grado de dificultad del que tiene el realizar la inferencia en distribuciones Normales Multivariadas.

Teorema III.1. Sea $\underline{y} \in \mathbb{R}^p$ un vector aleatorio con distribución esférica, entonces todas las marginales de \underline{y} de cualquier dimensión serán esféricas. En particular si $\underline{y} = [\underline{y}^{(1)'} \ \underline{y}^{(2)'}]'$, con $\underline{y}^{(1)} \in \mathbb{R}^m$ ($0 < m < p$) y $\underline{y}^{(2)} \in \mathbb{R}^n$ ($m + n = p$), entonces la función $g_m(t)$ que caracteriza a la distribución de $\underline{y}^{(1)}$ está dada por:

$$g_m(t) = \frac{\pi^{\frac{p-m}{2}}}{\Gamma(\frac{p-m}{2})} \int_t^\infty (y-t)^{\frac{p-m}{2}-1} g(y) dy \quad (III.3)$$

donde $g(t)$ es la función que caracteriza a la distribución de \underline{y} .

Demostración: se encuentra en [Fang, Kotz y Ng, 1990, pag. 36].

Teorema III.2. La distribución conjunta de $\underline{x}_1, \dots, \underline{x}_n$ vectores aleatorios independientes tales que $\underline{x}_i \in \mathbb{R}^p$ con idéntica distribución elíptica $\underline{x}_i \sim EC_p(\underline{\mu}, \underline{\Sigma}, g)$ está dada por:

$$|\underline{\Sigma}|^{-\frac{n}{2}} \prod_{i=1}^n g[(\underline{x}_i - \underline{\mu})' \underline{\Sigma}^{-1} (\underline{x}_i - \underline{\mu})] \quad (III.4)$$

Demostración: es directa dado que los vectores son independientes, por lo cual su densidad conjunta es el producto de las densidades individuales.

Definición III.3. Un conjunto de n vectores $\underline{x}_1, \dots, \underline{x}_n$, $\underline{x}_i \in \mathbb{R}^p$ con distribución multivariada pueden representarse por la matriz de dimensión $n \times p$ dada por:

$$X = (\underline{x}_1 \ \underline{x}_2 \ \dots \ \underline{x}_n)' \quad (III.5)$$

Proposición III.1. Sean $\underline{x}_1, \dots, \underline{x}_n$ vectores aleatorios independientes $\underline{x}_i \in \mathbb{R}^p$ con idéntica distribución Normal Multivariada $\underline{x}_i \sim N_p(\underline{\mu}, \underline{\Sigma})$, entonces la densidad de la matriz X dada por la **Definición III.1** es, con $\underline{1}' = (1, \dots, 1)$, $\underline{1} \in \mathbb{R}^n$:

$$\frac{1}{(2\pi)^{\frac{np}{2}} |\underline{I}_n \otimes \underline{\Sigma}|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} [\text{vec}(X') - \underline{1} \otimes \underline{\mu}]' [\underline{I}_n \otimes \underline{\Sigma}]^{-1} [\text{vec}(X') - \underline{1} \otimes \underline{\mu}] \right\} \quad (III.6)$$

Demostración:

En el caso de la distribución Normal Multivariada se tiene que:

$$g(t) = \frac{1}{(2\pi)^{\frac{p}{2}}} \exp\left\{-\frac{1}{2}t\right\}$$

y partir del **Teorema III.2** se tiene que la distribución conjunta de $\underline{x}_1, \dots, \underline{x}_n$ es:

$$\begin{aligned} & |\Sigma|^{-\frac{n}{2}} \prod_{i=1}^n \frac{1}{(2\pi)^{\frac{p}{2}}} \exp\left\{-\frac{1}{2}(\underline{x}_i - \underline{\mu})' \Sigma^{-1}(\underline{x}_i - \underline{\mu})\right\} \\ &= \frac{1}{(2\pi)^{\frac{np}{2}}} |\Sigma|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2} \sum_{i=1}^n (\underline{x}_i - \underline{\mu})' \Sigma^{-1}(\underline{x}_i - \underline{\mu})\right\} \\ &= \frac{1}{(2\pi)^{\frac{np}{2}} |I_n \otimes \Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2} [\text{vec}(X') - \underline{1} \otimes \underline{\mu}]' [I_n \otimes \Sigma]^{-1} [\text{vec}(X') - \underline{1} \otimes \underline{\mu}]\right\} \end{aligned}$$

QED

Estudiar la densidad de la matriz X es equivalente a estudiar a la densidad del vector $\text{vec}(X')$. La **Proposición III.1** plantea un modelo alternativo para definir la distribución de la matriz X cuando los vectores que la conforman tienen distribución elíptica.

Proposición III.2. Sean $\underline{x}_1, \dots, \underline{x}_n$ vectores aleatorios $\underline{x}_i \in \mathbb{R}^p$ no correlacionados con idéntica distribución elíptica $\underline{x}_i \sim EC_p(\underline{\mu}, \Sigma, g)$, que definen a la matriz X , entonces la densidad de X es:

$$\begin{aligned} & |I_n \otimes \Sigma|^{-\frac{1}{2}} g^* \left([\text{vec}(X') - \underline{1} \otimes \underline{\mu}]' [I_n \otimes \Sigma]^{-1} [\text{vec}(X') - \underline{1} \otimes \underline{\mu}] \right) \\ &= |\Sigma|^{-\frac{n}{2}} g^* \left(\sum_{i=1}^n (\underline{x}_i - \underline{\mu})' \Sigma^{-1}(\underline{x}_i - \underline{\mu}) \right) \end{aligned} \tag{III.7}$$

para alguna función $g^*(t)$ con la misma forma funcional que la función $g(t)$ que caracteriza a la distribución elíptica de los vectores $\underline{x}_1, \dots, \underline{x}_n$.

Demostración: se encuentra en [Fang y Anderson, 1990a, pag. v].

El resultado de la **Proposición III.2** plantea que, para analizar estadísticamente una muestra aleatoria, proveniente de una distribución elíptica, se puede partir de la distribución

de la matriz X , en vez de partir de la distribución conjunta definida en el **Teorema III.2**, esto debido a que la independencia implica la no correlación de los individuos. Dependiendo de la aplicación que se esté considerando, es probable que una de las posibilidades permita encontrar resultados de manera más simple. Nótese a partir de la **Proposición III.1**, que para el caso de la distribución Normal Multivariada ambas alternativas de análisis son equivalentes.

Dado que la distribución de la matriz X es la misma que la distribución del vector $vec(X')$, la **Proposición III.2**, deja en claro que $vec(X')$ entonces tendrá distribución elíptica, para la cual existirá V^* definida por $V^* = R^{*2} = \underline{y}^* \underline{y}^{*'} = \|\underline{y}^*\|^2$ con $E(vec(X')) = \underline{1} \otimes \underline{\mu}$, y $Var(vec(X')) = \frac{E(V^*)}{np} [I_n \otimes \underline{\Sigma}]$, donde V^* tendrá la distribución:

$$f(v^*) = \frac{\pi^{\frac{np}{2}}}{\Gamma(\frac{np}{2})} v^{*(\frac{np}{2})-1} g^*(v) \quad (III.8)$$

Además, dado que en una muestra aleatoria las columnas de X' serán independientes, es claro que los bloques diagonales que conforman a la matriz de varianzas y covarianzas de $vec(X')$ son todos iguales y corresponden a la varianza de \underline{x} con la distribución elíptica correspondiente a cada individuo de la muestra, esto es, $\underline{x} \sim EC_p(\underline{\mu}, \underline{\Sigma}, g)$. Esto implica que $\frac{E(V^*)}{np} = \frac{E(V)}{p}$.

El resultado, que puede reescribirse como $E(V^*) = nE(V)$, puede simplificar la identificación de $g^*(t)$ en aquellos casos en los que, multiplicar por n a la esperanza de V , equivale simplemente a multiplicar por n un parámetro de la distribución $f(v)$, como en el caso de la t de Student multivariada o de la normal multivariada. Sin embargo, hay ocasiones en que esto no ocurre, y para esos casos se puede obtener $g^*(t)$ a partir del **Teorema III.1**. Esto se ejemplificará más adelante para el caso de la distribución Pearson Tipo VII.

Teorema III.3. Sean $\underline{x}_1, \dots, \underline{x}_n$ vectores aleatorios independientes con idéntica distribución elíptica $\underline{x}_i \sim EC_p(\underline{\mu}, \underline{\Sigma}, g)$, entonces la distribución de la matriz X está dada por:

$$f(X) = |\underline{\Sigma}|^{-\frac{n}{2}} g^*(tr \underline{\Sigma}^{-1} G) \quad (III.9)$$

donde $G = \sum_{i=1}^n (\underline{x}_i - \underline{\mu})(\underline{x}_i - \underline{\mu})'$ y $g^*(t)$ corresponde a la función definida en la **Proposición III.2**.

Demostración: se encuentra en [Fang y Anderson, 1990a, pag. 9].

Teorema III.4. Sean $\underline{x}_1, \dots, \underline{x}_n$ vectores aleatorios independientes con idéntica distribución elíptica $\underline{x}_i \sim EC_p(\underline{\mu}, \underline{\Sigma}, g)$, donde la función $g^*(t)$ definida en la **Proposición III.2** sea decreciente y diferenciable, entonces el estimador máximo verosímil de $\underline{\Sigma}$ denotado como $\widehat{\underline{\Sigma}}$ es:

$$\widehat{\underline{\Sigma}} = \delta_{\max(g^*)} W \quad (III.10)$$

donde $\delta_{\max(g^*)}$ es el valor que maximiza la función $f(\delta) = \delta^{-\frac{np}{2}} g^*\left(\frac{p}{\delta}\right)$ siendo n el tamaño de muestra y donde $W = \sum_{i=1}^n (\underline{x}_i - \bar{\underline{x}})(\underline{x}_i - \bar{\underline{x}})'$.

Demostración: se encuentra en [Fang y Anderson, 1990b, pag. 205].

Lema III.1. El valor $\delta_{\max(g^*)}$ que maximiza la función $f(\delta) = \delta^{-\frac{np}{2}} g^*\left(\frac{p}{\delta}\right)$ bajo las condiciones del **Teorema III.4** es solución de la ecuación:

$$\delta_{\max(g^*)} = \left(\frac{-2}{n}\right) \frac{g^{*'}\left(\frac{p}{\delta_{\max(g^*)}}\right)}{g^*\left(\frac{p}{\delta_{\max(g^*)}}\right)} \quad (III.11)$$

Demostración: se encuentra en [Fang y Anderson, 1990b, pag. 204].

Corolario III.1. Sean $\underline{x}_1, \dots, \underline{x}_n$ vectores aleatorios independientes con idéntica distribución elíptica $\underline{x}_i \sim EC_p(\underline{\mu}, \underline{\Sigma}, g)$, donde la función $g^*(t)$ definida en la **Proposición III.2** sea decreciente y diferenciable, entonces el estimador máximo verosímil de $Var(\underline{x}) = \underline{\Psi}$ denotado como $\widehat{\underline{\Psi}}$ es:

$$\begin{aligned} \widehat{\underline{\Psi}} &= \frac{E(R^2)}{p} \delta_{\max(g^*)} W \\ &= \frac{E(R^{*2})}{np} \delta_{\max(g^*)} W \end{aligned} \quad (III.12)$$

Demostración: se deriva de manera directa de la propiedad de invarianza de los estimadores máximo verosímiles.

Teorema III.5. Sean $\underline{x}_1, \dots, \underline{x}_n$ vectores aleatorios independientes con idéntica distribución elíptica, y suponiendo sin pérdida de generalidad que $\underline{\mu} = \underline{0}$, entonces la densidad de la matriz $W = \sum_{i=1}^n (\underline{x}_i - \bar{\underline{x}})(\underline{x}_i - \bar{\underline{x}})'$ es:

$$f(W) = \frac{\pi^{\frac{np}{2} - \frac{p(p-1)}{4}}}{\prod_{\alpha=1}^p \Gamma[\frac{n-\alpha+1}{2}]} |W|^{\frac{n-p-1}{2}} |\Sigma|^{\frac{-n}{2}} g^*(tr(\Sigma^{-1}W)) \quad (III.13)$$

Demostración: se encuentra en [Fang y Anderson, 1990a, pag. 12].

III.3 Ejemplo (Distribución Pearson Tipo VII)

Definición III.4. Un vector aleatorio $\underline{x} \in \mathbb{R}^p$ tiene distribución Pearson Tipo VII denotada como $\underline{x} \sim PP_p(\underline{\mu}, \gamma, m, \Sigma, g)$ si la función $g(t)$ que la caracteriza está dada por:

$$g(t) = \frac{\Gamma(m) \left(\frac{1}{\gamma}\right)^{\frac{p}{2}}}{\pi^{\frac{p}{2}} \Gamma(m - \frac{p}{2})} \left(1 + \frac{t}{\gamma}\right)^{-m} \quad (III.14)$$

con los siguientes parámetros:

- i) $\underline{\mu} \in \mathbb{R}$ parámetro de localización;
- ii) $\Sigma \in \mathbb{R}^p$ parámetro de escala;
- iii) $\gamma > 0$ grados de libertad;
- iv) $m > \frac{p}{2}$ parámetro de forma.

Algunos casos particulares o casos límite de esta distribución incluyen a las distribuciones t de Student multivariada, exponencial multivariada y normal multivariada, esto bajo valores específicos de los parámetros, los cuales permiten encontrar un mejor ajuste de los datos a la distribución.

Proposición III.3. Sean $\underline{x}_1, \dots, \underline{x}_n$ vectores aleatorios independientes teniendo idéntica distribución Pearson Tipo VII $\underline{x}_i \sim PP_p(\underline{\mu}, \gamma, m, \Sigma, g)$, entonces la distribución de la matriz X está dada por:

$$f(X) = |\Sigma|^{-\frac{n}{2}} g^*(tr \Sigma^{-1} X) \quad (III.15)$$

donde

$$g^*(t) = \frac{\Gamma(\frac{2m+np-p}{2}) \left(\frac{1}{\gamma}\right)^{\frac{np}{2}}}{\pi^{\frac{np}{2}} \Gamma(m - \frac{p}{2})} \left(1 + \frac{t}{\gamma}\right)^{-\frac{2m+np-p}{2}} \quad (III.16)$$

Demostración:

La ***Proposición III.2*** establece que $g^*(t)$ es la función que caracteriza al vector $vec(X')$, siendo $g^*(t)$ de la misma familia que $g(t)$, por lo cual en el caso de la distribución Pearson Tipo VII se tiene que:

$$g^*(t) = \frac{\Gamma(m^*) \left(\frac{1}{\gamma^*}\right)^{\frac{p^*}{2}}}{\pi^{\frac{p^*}{2}} \Gamma(m^* - \frac{p^*}{2})} \left(1 + \frac{t}{\gamma^*}\right)^{-m^*}$$

Esto implica que $vec(X') \sim PP_{p^*}(\mathbf{1} \otimes \underline{\mu}, \gamma^*, m^*, I_n \otimes \Sigma, g^*)$, y entonces la distribución del vector de dimensión $np \times 1$ definido por $\underline{y} = A^{-1}(vec(X') - \mathbf{1} \otimes \underline{\mu})$, con $A'A = I_n \otimes \Sigma$, es $\underline{y} \sim PP_{p^*}(\mathbf{0}, \gamma^*, m^*, I_{p^*}, g^*)$. Si se particiona a \underline{y} como $\underline{y} = [\underline{y}^{(1)'} \ \underline{y}^{(2)'}]'$ con $\underline{y}^{(1)}$ de dimensión $p \times 1$, es claro que la distribución de $\underline{y}^{(1)}$ debe corresponder a $y^{(1)} \sim PP_{p^*}(\mathbf{0}, \gamma, m, I_{p^*}, g)$. Para encontrar los parámetros de $g^*(t)$ se puede utilizar entonces el ***Teorema III.1*** que permite relacionar $g(t)$ con $g^*(t)$ de la siguiente forma:

$$g(t) = \frac{\pi^{\frac{np-p}{2}}}{\Gamma(\frac{np-p}{2})} \int_t^\infty (y-t)^{\frac{np-p}{2}-1} g^*(y) dy$$

Se requiere entonces desarrollar el factor derecho de la ecuación anterior, para llegar a una forma similar a $g(t)$ que de antemano se sabe es:

$$g(t) = \frac{\Gamma(m) \left(\frac{1}{\gamma}\right)^{\frac{p}{2}}}{\pi^{\frac{p}{2}} \Gamma(m - \frac{p}{2})} \left(1 + \frac{t}{\gamma}\right)^{-m}$$

Desarrollando el factor referido, con el cambio de variable dado por $s = y - t$, de donde $ds = dy$, con s en el intervalo $(0, \infty)$ se tiene:

$$g(t) = \frac{\pi^{\frac{np-p}{2}}}{\Gamma(\frac{np-p}{2})} \frac{\Gamma(m^*)}{\pi^{\frac{p^*}{2}} \Gamma(m^* - \frac{p^*}{2})} \left(\frac{1}{\gamma^*}\right)^{\frac{p^*}{2}} \int_0^\infty (s)^{\frac{np-p}{2}-1} \left(1 + \frac{s+t}{\gamma^*}\right)^{-m^*} ds$$

Es sencillo probar que:

$$\left(1 + \frac{s+t}{\gamma^*}\right)^{-m^*} = \left(\frac{\gamma^* + t}{\gamma^*}\right)^{-m^*} \left(1 + \frac{s}{\gamma^* + t}\right)^{-m^*}$$

Con lo cual:

$$g(t) = \frac{\pi^{\frac{np-p}{2} - \frac{\rho^*}{2}} \Gamma(m^*)}{\Gamma(\frac{np-p}{2}) \Gamma(m^* - \frac{\rho^*}{2})} \left(\frac{1}{\gamma^*}\right)^{\frac{\rho^*}{2}} \int_0^\infty (s)^{\frac{np-p}{2} - 1} \left(\frac{\gamma^* + t}{\gamma^*}\right)^{-m^*} \left(1 + \frac{s}{\gamma^* + t}\right)^{-m^*} ds$$

Sean $a = np - p$ y $b = 2m^* - a$ de modo que $a + b = 2m^*$. Se define un nuevo cambio de variable dado por: $w = \left(\frac{s}{\gamma^* + t}\right) \left(\frac{b}{a}\right)$ de modo que $s = w(\gamma^* + t)\frac{a}{b}$ y $ds = (\gamma^* + t)\frac{a}{b} dw$, con w en el intervalo $(0, \infty)$, de donde:

$$g(t) = \frac{\pi^{\frac{np-p}{2} - \frac{\rho^*}{2}} \Gamma(m^*)}{\Gamma(\frac{np-p}{2}) \Gamma(m^* - \frac{\rho^*}{2})} \left(\frac{1}{\gamma^*}\right)^{\frac{\rho^*}{2}} \left(\frac{\gamma^* + t}{\gamma^*}\right)^{-m^*} (\gamma^* + t)^{\frac{np-p}{2}} \int_0^\infty (w)^{\frac{a}{2} - 1} \left(\frac{a}{b}\right)^{\frac{a}{2}} \left(1 + \frac{a}{b}w\right)^{-\frac{a+b}{2}} dw$$

Si se identifica al término en la integral como una distribución $F(a, b)$ con ausencia de algunas constantes, se puede completar de modo que se obtiene:

$$\begin{aligned} g(t) &= \frac{\pi^{\frac{np-p}{2} - \frac{\rho^*}{2}} \Gamma(m^*)}{\Gamma(\frac{np-p}{2}) \Gamma(m^* - \frac{\rho^*}{2})} \left(\frac{1}{\gamma^*}\right)^{\frac{\rho^*}{2}} \left(\frac{\gamma^* + t}{\gamma^*}\right)^{-m^*} (\gamma^* + t)^{\frac{np-p}{2}} \frac{\Gamma\left(\frac{a}{2}\right) \Gamma\left(\frac{b}{2}\right)}{\Gamma\left(\frac{a+b}{2}\right)} \\ &= \frac{\pi^{\frac{np-p}{2} - \frac{\rho^*}{2}} \Gamma(m^*)}{\Gamma(\frac{np-p}{2}) \Gamma(m^* - \frac{\rho^*}{2})} \left(\frac{1}{\gamma^*}\right)^{\frac{\rho^*}{2}} \left(\frac{\gamma^* + t}{\gamma^*}\right)^{-m^*} (\gamma^* + t)^{\frac{np-p}{2}} \frac{\Gamma\left(\frac{np-p}{2}\right) \Gamma\left(\frac{2m^* - np + p}{2}\right)}{\Gamma(m^*)} \\ &= \frac{\pi^{\frac{np-p}{2} - \frac{\rho^*}{2}} \Gamma\left(\frac{2m^* - np + p}{2}\right)}{\Gamma(m^* - \frac{\rho^*}{2})} \left(\frac{1}{\gamma^*}\right)^{\frac{\rho^*}{2} - \frac{np-p}{2}} \left(1 + \frac{t}{\gamma^*}\right)^{-m^* + \frac{np-p}{2}} \end{aligned}$$

A partir de esta ecuación se puede determinar que:

$$\begin{aligned} \frac{np - p - \rho^*}{2} &= -\frac{\rho}{2} \Rightarrow \rho^* = np \\ \frac{2m^* - np + p}{2} &= m \Rightarrow m^* = \frac{2m + np - p}{2} \\ \gamma^* &= \gamma \end{aligned}$$

De donde:

$$g^*(t) = \frac{\Gamma\left(\frac{2m + np - p}{2}\right) \left(\frac{1}{\gamma}\right)^{\frac{np}{2}}}{\pi^{\frac{np}{2}} \Gamma(m - \frac{\rho}{2})} \left(1 + \frac{t}{\gamma}\right)^{-\frac{2m + np - p}{2}}$$

QED

Proposición III.4. Sean $\underline{x}_1, \dots, \underline{x}_n$ vectores aleatorios independientes teniendo idéntica distribución Pearson Tipo VII $\underline{x}_j \sim PP_p(\underline{\mu}, \gamma, m, \Sigma, g)$, donde $m > \frac{p+2}{2}$, la matriz de varianzas y covarianzas Ψ está dada por:

$$\Psi = \frac{\gamma \Sigma}{2m - p - 2} \quad (III.17)$$

Demostración:

Por la **Definición III.1**, la distribución $f(v)$ es:

$$f(v) = \frac{\Gamma(m)}{\Gamma(m - \frac{p}{2})\Gamma(\frac{p}{2})} \left(\frac{1}{\gamma}\right) \left(\frac{v}{\gamma}\right)^{\frac{p}{2}-1} \left(1 + \frac{v}{\gamma}\right)^{-m}$$

Si se definen $a = p$ y $b = 2m - p$ de modo que $a + b = 2m$ y se propone la transformación $w = \left(\frac{b}{a}\right) \left(\frac{v}{\gamma}\right)$ de modo que $v = \frac{a}{b} w \gamma$, siendo el jacobiano de la transformación igual a $\frac{a}{b} \gamma$, entonces la distribución de w es:

$$f(w) = \frac{\Gamma(\frac{a+b}{2})}{\Gamma(\frac{a}{2})\Gamma(\frac{b}{2})} \left(\frac{a}{b}\right)^{\frac{a}{2}} w^{\frac{a}{2}-1} \left(1 + w \left(\frac{a}{b}\right)\right)^{-\frac{a+b}{2}}$$

Que corresponde a una distribución $F(a, b)$, y por lo cual $E(w) = \frac{b}{b-2} = \frac{2m-p}{2m-p-2}$ de modo que:

$$E(v) = \frac{a}{b} E(w) \gamma = \left(\frac{p}{2m-p}\right) \left(\frac{2m-p}{2m-p-2}\right) \gamma = \frac{\gamma p}{2m-p-2}$$

Dado que $\Psi = \frac{E(v)}{p} \Sigma$, entonces se tiene que:

$$\Psi = \frac{\gamma \Sigma}{2m - p - 2}$$

QED

Proposición III.5. Sean $\underline{x}_1, \dots, \underline{x}_n$ vectores aleatorios independientes teniendo idéntica distribución Pearson Tipo VII $\underline{x}_j \sim PP_p(\underline{\mu}, \gamma, m, \Sigma, g)$, los estimadores máximo verosímiles de la matriz parámetro Σ y de la matriz de varianzas y covarianzas Ψ están dados por:

$$\hat{\Sigma} = \left(\frac{2m-p}{\gamma}\right) \frac{W}{n} \quad y \quad \hat{\Psi} = \left(\frac{2m-p}{2m-p-2}\right) \frac{W}{n} \quad (III.18)$$

donde $m > \frac{p+2}{2}$.

Demostración:

La función $g^*(t)$ que se encontró en la **Proposición III.3** para la distribución Pearson tipo VII es decreciente y diferenciable, por tanto, por el **Teorema III.4**, el estimador máximo verosímil de Σ es $\widehat{\Sigma} = \delta_{\max(g^*)}W$. Para obtener $\delta_{\max(g^*)}$ se utiliza el **Lema III.1**, para ello se obtiene el cociente

$$\frac{g^{*\prime}(t)}{g^*(t)}$$

Dado que:

$$g^*(t) = \frac{\Gamma(m^*) \left(\frac{1}{\gamma^*}\right)^{\frac{p^*}{2}}}{\pi^{\frac{p^*}{2}} \Gamma(m^* - \frac{p^*}{2})} \left(1 + \frac{t}{\gamma^*}\right)^{-m^*}$$

con

$$\begin{aligned} p^* &= np \\ m^* &= \frac{2m + np - p}{2} \\ \gamma^* &= \gamma \end{aligned}$$

entonces

$$g^{*\prime}(t) = \frac{\Gamma(m^*) \left(\frac{1}{\gamma^*}\right)^{\frac{p^*}{2}}}{\pi^{\frac{p^*}{2}} \Gamma(m^* - \frac{p^*}{2})} (-m^*) \left(1 + \frac{t}{\gamma^*}\right)^{-m^*-1} \left(\frac{1}{\gamma^*}\right)$$

y por tanto

$$\frac{g^{*\prime}(t)}{g^*(t)} = \frac{-m^*}{\gamma^* + t}$$

De este modo, para obtener $\delta_{max(g)}$ el **Lema III.1** establece la necesidad de despejar de la siguiente ecuación:

$$\begin{aligned}\delta_{max(g^*)} &= \left(\frac{-2}{n}\right) \frac{g^{*'}\left(\frac{p}{\delta_{max(g^*)}}\right)}{g^*\left(\frac{p}{\delta_{max(g^*)}}\right)} \\ &= \left(\frac{-2}{n}\right) \left(\frac{-m^*}{\gamma + \frac{p}{\delta_{max(g^*)}}}\right)\end{aligned}$$

de donde:

$$\begin{aligned}\delta_{max(g^*)} &= \frac{2m^* - np}{n\gamma} \\ &= \frac{2m + np - p - np}{n\gamma} \\ &= \frac{2m - p}{n\gamma}\end{aligned}$$

y por tanto

$$\hat{\Sigma} = \left(\frac{2m - p}{\gamma}\right) \frac{W}{n}$$

finalmente, por el **Corolario III.1** y por el siguiente resultado, previamente probado

$$E(v) = \frac{\gamma p}{2m - p - 2}$$

se tiene que

$$\hat{\Psi} = \left(\frac{2m - p}{2m - p - 2}\right) \frac{W}{n}$$

QED

III.4 Simulación de Vectores con Distribución Elíptica

Se han estudiado diversas formas de obtener datos simulados provenientes de distribuciones elípticas, y son tres métodos los más recurridos. El primero consiste en obtener distribuciones condicionales (lo cual es simple para las distribuciones elípticas) y a partir de dichas

distribuciones utilizar algoritmos de simulación de Monte Carlo vía Cadenas de Markov, con precisión empíricamente comprobada (Johnson, 1987). El segundo tipo de métodos son los basados en mezclas de escalas, consistentes en encontrar las mezclas de distribuciones más simples de simular, que generan a su vez a las distribuciones elípticas deseadas. Este segundo tipo tiene la desventaja de limitar los valores de los parámetros a ciertos rangos, por lo cual no es general (Haro-López, 1998). Una tercer forma de simular consiste en partir de lo que se ha denominado representación de Cambanis para las distribuciones elípticas. Esta metodología es de carácter general pero poco conocida y aplicada, por lo que se presenta a continuación (Johnson, 1987).

En la **Sección III.1** se establece que para un vector aleatorio $\underline{x} \in \mathbb{R}^p$ con distribución elíptica denotada por $\underline{x} \sim EC_p(\underline{\mu}, \Sigma, g)$ se tiene un vector aleatorio \underline{y} con distribución esférica (elíptica estandarizada) correspondiente a la distribución elíptica de \underline{x} de modo que, para $AA' = \Sigma$, la distribución de \underline{x} es la misma que la de $\underline{\mu} + RA\underline{u}$ siendo \underline{u} un vector uniforme sobre la esfera de dimensión p , donde la distribución de $V = R^2$ es:

$$f(v) = \frac{\pi^{\frac{p}{2}}}{\Gamma(\frac{p}{2})} v^{(\frac{p}{2})-1} g(v) \quad (III.19)$$

Este resultado, que se conoce como representación de Cambanis, permite definir un algoritmo de simulación para datos provenientes de la distribución elíptica $\underline{x} \sim EC_p(\underline{\mu}, \Sigma, g)$ de la siguiente forma:

Proposición III.6. Un algoritmo para generar un vector $\underline{x} \in \mathbb{R}^p$ con distribución elíptica denotada por $\underline{x} \sim EC_p(\underline{\mu}, \Sigma, g)$ está dado por:

- a)** Generar un vector $\underline{u} \in \mathbb{R}^p$ uniforme sobre la esfera unitaria de dimensión p .
- b)** Generar un valor R .
- c)** Obtener A tal que $AA' = \Sigma$
- d)** Calcular $\underline{x} = \underline{\mu} + RA\underline{u}$

Demostración: es directa a partir de los resultados previamente presentados.

Para la generación de los vectores referidos en el inciso **a)** se puede utilizar el algoritmo que se describe a continuación.

Proposición III.7. Un algoritmo para generar el vector $\underline{u} \in \mathbb{R}^p$ uniforme en la esfera de dimensión p requerido en el inciso a) de la **Proposición III.6** consiste en:

i) Generar p valores U_1, \dots, U_n provenientes de p variables aleatorias independientes con la misma distribución uniforme en el intervalo $[-1, 1]$

ii) Calcular $S^2 = U_1^2 + \dots + U_n^2$, y repetir el paso a) hasta que $S \leq 1$.

iii) Generar $\underline{u} = \left(\frac{U_1}{S}, \dots, \frac{U_n}{S} \right)$

Demostración: este algoritmo se considera como de aceptación-rechazo, y se basa en el hecho de que la esfera unitaria de dimensión p está inscrita totalmente en el hipercubo con lados de longitud 2 centrado en el origen, a partir de lo cual se plantea generar vectores dentro del hipercubo, y esperar hasta obtener uno que esté dentro de la esfera, y posteriormente normalizar dicho vector para que satisfaga la condición de estar sobre la esfera unitaria (Devroye, 1986).

Para obtener el valor requerido en el inciso b) de la **Proposición III.6** se plantea la posibilidad de utilizar algoritmos de aceptación-rechazo, partiendo de la distribución de R , o en su caso, de una función de R sencilla de simular por medio de los algoritmos referidos. Se ejemplificarán los casos de las distribuciones Normal Multivariada y Pearson Tipo VII.

Lema III.2. Sea $\underline{x} \in \mathbb{R}^p$ un vector aleatorio con distribución Normal Multivariada, entonces $V = R^2$ tiene distribución χ^2 con p grados de libertad.

Demostración:

Para una distribución Normal Multivariada se tiene que:

$$g(v) = (2\pi)^{-\frac{p}{2}} e^{-\frac{1}{2}v}$$

de donde

$$f(v) = \frac{\left(\frac{1}{2}\right)^{\frac{p}{2}}}{\Gamma\left(\frac{p}{2}\right)} v^{\frac{p}{2}-1} e^{-\frac{1}{2}v}$$

función que corresponde a una distribución χ^2 con p grados de libertad.

QED

Lema III.3. Sea $W \in \mathbb{R}^+$ una variable aleatoria con distribución χ^2 con p grados de libertad, entonces $Y = \frac{1}{2}W$ tiene distribución $Gamma(\frac{p}{2}, 1)$.

Demostración:

Si $Y = \frac{1}{2}W$ entonces $W = 2Y$ y por tanto el jacobiano de la transformación vale 2, por lo cual la distribución de Y es:

$$\begin{aligned} f(y) &= |2| \frac{\left(\frac{1}{2}\right)^{\frac{p}{2}}}{\Gamma\left(\frac{p}{2}\right)} (2y)^{\frac{p}{2}-1} e^{-\frac{1}{2}2y} \\ &= \frac{1}{\Gamma\left(\frac{p}{2}\right)} y^{\frac{p}{2}-1} e^{-y} \end{aligned}$$

función que corresponde a una distribución $Gamma(\frac{p}{2}, 1)$.

QED

Corolario III.2. Sea $\underline{x} \in \mathbb{R}^p$ un vector aleatorio con distribución Normal Multivariada, entonces $Y = \frac{R^2}{2}$ tiene distribución $Gamma(\frac{p}{2}, 1)$.

Demostración: es directa a partir de los **Lemas III.2 y III.3**.

Proposición III.8. Un algoritmo para generar R correspondiente a una distribución Normal Multivariada consiste en:

i) Generar un valor y de una distribución $Gamma(\frac{p}{2}, 1)$

ii) Calcular $R = (2y)^{\frac{1}{2}}$.

Demostración: es directa a partir del **Corolario III.2**.

Para implementar de manera práctica este algoritmo se puede hacer uso de algoritmos de aceptación y rechazo de probada eficiencia para la generación de valores provenientes de la distribución $Gamma(\frac{p}{2}, 1)$, como por ejemplo el Algoritmo XG de rechazo de Best para variables aleatorias $Gamma$ (Devroye, 1986).

Lema III.4. Sea $\underline{x} \in \mathbb{R}^p$ un vector aleatorio con distribución Pearson tipo VII con parámetros m y γ , entonces $V = R^2$ tiene la siguiente distribución:

$$f(v) = \frac{\Gamma(m)}{\Gamma(m - \frac{p}{2}) \Gamma(\frac{p}{2})} \left(\frac{1}{\gamma}\right)^{\frac{p}{2}} v^{\frac{p}{2}-1} \left(1 + \frac{v}{\gamma}\right)^{-m} \quad (III.20)$$

Demostración:

Para una distribución Pearson tipo VII con parámetros m y γ se tiene que:

$$g(v) = \frac{\Gamma(m) \left(\frac{1}{\gamma}\right)^{\frac{p}{2}}}{\pi^{\frac{p}{2}} \Gamma(m - \frac{p}{2})} \left(1 + \frac{v}{\gamma}\right)^{-m}$$

de donde

$$f(v) = \frac{\Gamma(m)}{\Gamma(m - \frac{p}{2}) \Gamma(\frac{p}{2})} \left(\frac{1}{\gamma}\right)^{\frac{p}{2}} v^{\frac{p}{2}-1} \left(1 + \frac{v}{\gamma}\right)^{-m}$$

función que corresponde a la enunciada en el lema.

QED

Lema III.5. Sea $W \in \mathbb{R}^+$ una variable aleatoria con distribución:

$$f(w) = \frac{\Gamma(m)}{\Gamma(m - \frac{p}{2}) \Gamma(\frac{p}{2})} \left(\frac{1}{\gamma}\right)^{\frac{p}{2}} w^{\frac{p}{2}-1} \left(1 + \frac{w}{\gamma}\right)^{-m} \quad (III.21)$$

entonces la distribución de Y dada por la transformación $Y = \frac{W}{\gamma} \left(\frac{2m-p}{p}\right)$ es $F(p, 2m - p)$.

Demostración:

Si $Y = \frac{W}{\gamma} \left(\frac{2m-p}{\rho} \right)$ entonces $W = Y\gamma \left(\frac{\rho}{2m-p} \right)$ y por tanto el jacobiano de la transformación vale $\gamma \left(\frac{\rho}{2m-p} \right)$, por lo cual la distribución de Y es:

$$\begin{aligned} f(y) &= \left| \gamma \left(\frac{\rho}{2m-p} \right) \right| \frac{\Gamma(m)}{\Gamma\left(m - \frac{\rho}{2}\right) \Gamma\left(\frac{\rho}{2}\right)} \left(\frac{1}{\gamma} \right)^{\frac{\rho}{2}} \left(y\gamma \left(\frac{\rho}{2m-p} \right) \right)^{\frac{\rho}{2}-1} \left(1 + \frac{\left(y\gamma \left(\frac{\rho}{2m-p} \right) \right)}{\gamma} \right)^{-m} \\ &= \frac{\Gamma(m)}{\Gamma\left(m - \frac{\rho}{2}\right) \Gamma\left(\frac{\rho}{2}\right)} \left(\frac{\rho}{2m-p} \right)^{\frac{\rho}{2}} y^{\frac{\rho}{2}-1} \left(1 + y \left(\frac{\rho}{2m-p} \right) \right)^{-m} \\ &= \frac{\Gamma\left(\frac{\rho+(2m-p)}{2}\right)}{\Gamma\left(\frac{\rho}{2}\right) \Gamma\left(\frac{2m-p}{2}\right)} \left(\frac{\rho}{2m-p} \right)^{\frac{\rho}{2}} y^{\frac{\rho}{2}-1} \left(1 + y \left(\frac{\rho}{2m-p} \right) \right)^{-\frac{\rho+(2m-p)}{2}} \end{aligned}$$

función que corresponde a una distribución $F(\rho, 2m - \rho)$.

QED

Corolario III.3. Sea $\underline{x} \in \mathbb{R}^p$ un vector aleatorio con distribución Pearson tipo VII con parámetros m y γ , entonces $Y = \frac{R^2}{\gamma} \left(\frac{2m-p}{\rho} \right)$ tiene distribución $F(\rho, 2m - \rho)$.

Demostración: es directa a partir de los **Lemas III.4 y III.5**.

Este resultado implica que para generar valores para R se pueden generar valores Y de una distribución $F(\rho, 2m - \rho)$, que van a corresponder a $Y = \frac{R^2}{\gamma} \left(\frac{2m-p}{\rho} \right)$, por lo cual los valores R serán:

$$R = \left(Y\gamma \left(\frac{\rho}{2m-p} \right) \right)^{\frac{1}{2}} \tag{III.22}$$

Un resultado bien conocido establece que si S_1 se distribuye χ^2 con n_1 grados de libertad, y S_2 se distribuye χ^2 con n_2 grados de libertad, con S_1 y S_2 independientes, entonces:

$$Y = \frac{\left(\frac{S_1}{n_1} \right)}{\left(\frac{S_2}{n_2} \right)}$$

tendrá distribución $F(n_1, n_2)$. Este resultado permite generar valores de Y a partir de valores generados de S_1 y S_2 independientes.

Proposición III.9. Un algoritmo para generar el valor de R correspondiente a una distribución Pearson Tipo VII con parámetros m y γ consiste en:

i) Generar un valor W_1 de una distribución $Gamma(\frac{p}{2}, 1)$.

ii) Generar un valor W_2 de una distribución $Gamma(\frac{2m-p}{2}, 1)$.

iii) Calcular $S_1 = 2W_1$ y $S_2 = 2W_2$.

iv) Calcular

$$Y = \frac{\frac{S_1}{p}}{\frac{S_2}{(2m-p)}}$$

v) Calcular $R = \left(Y \gamma \left(\frac{p}{2m-p} \right) \right)^{\frac{1}{2}}$.

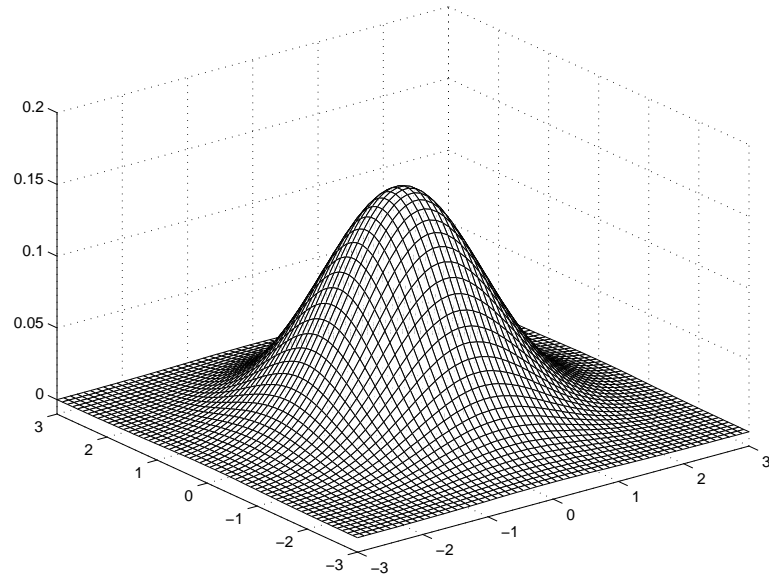
Demostración: es directa a partir del **Lema III.3**, del **Lema III.5** y del resultado conocido que relaciona las distribuciones χ^2 con la distribución F .

Finalmente, cabe señalar que una forma de calcular la matriz A referida en el inciso **c)** de la **Proposición III.6** consiste en aplicar la descomposición SVD a la matriz Σ , que al ser simétrica generará un producto de la forma: UDU' . Al obtenerse la descomposición de Cholesky de D en la forma: $E'E = D$ se puede generar A como $A = UE'$ (Golub, et. al., 1989).

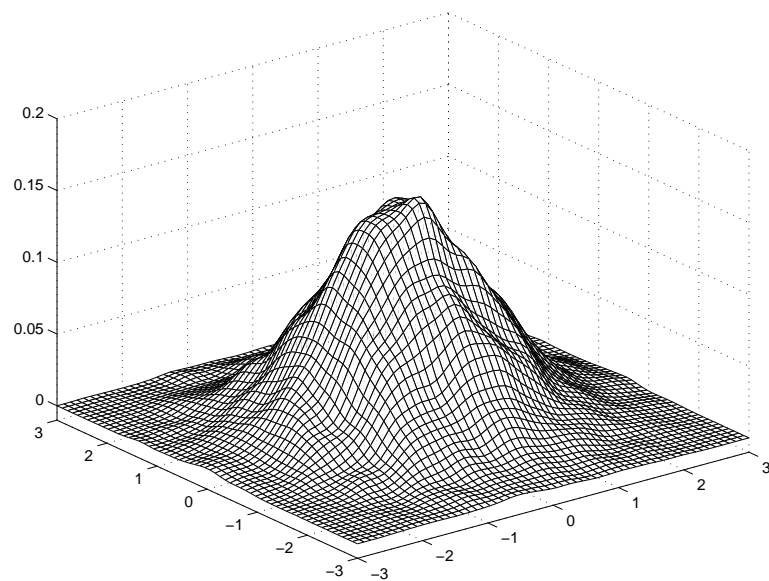
A continuación se presentan las gráficas de dos distribuciones elípticas para las que se utilizó el método de simulación descrito en esta sección. Se presentan a manera de comparación las gráficas de las distribuciones reales con las distribuciones empíricas de los datos simulados obtenidas por medio de estimación por kernel bivariada. Se utilizó, para la generación de los datos, el programa en lenguaje *Fortran* enunciado en el Apéndice, y para la generación de las gráficas de estimación por kernel bivariada, la función *kde2d* disponible en el paquete *R*.

Tanto en el caso de la distribución Normal bivariada, como en el de la Pearson Tipo VII, se observa una similitud notoria entre las gráficas teórica y empírica, aunque es claro que no son completamente iguales. Esta comparación visual se presenta para mostrar la viabilidad del uso del método de simulación previamente descrito.

Grafica III.1. Distribución Normal Bivariada con media $\underline{\mu} = (0, 0)$ y matriz de varianzas y covarianzas $\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

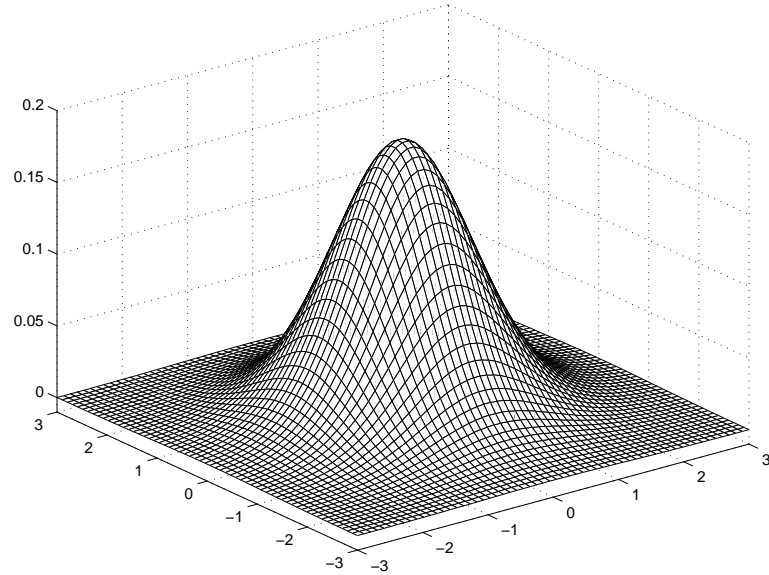


Grafica III.2. Distribución Normal Bivariada simulada con media $\underline{\mu} = (0, 0)$ y matriz de varianzas y covarianzas $\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ ($n = 10,000$)



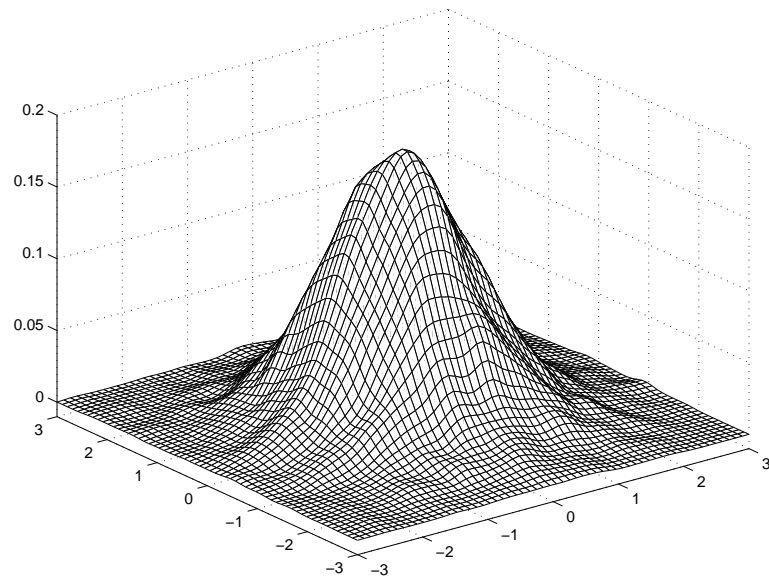
Grafica III.3. Distribución Pearson Tipo VII Bivariada con media $\underline{\mu} = (0, 0)$ y parámetros

$$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \gamma = 15 \text{ y } m = 10$$



Grafica III.4. Distribución Pearson Tipo VII Bivariada simulada con media $\underline{\mu} = (0, 0)$ y

parámetros $\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \gamma = 15$ y $m = 10$ ($n = 10,000$)



Se presentan a continuación las comparaciones entre las medias poblacionales y las correspondientes a los datos simulados de las distribuciones previamente graficadas. En el caso de la distribución Normal bivariada se tiene que la media poblacional es:

$$E(\underline{x}) = (0, 0)$$

mientras que la media de los datos simulados es:

$$\bar{\underline{x}} = (0.012, 0.004)$$

La matriz de varianzas y covarianzas poblacional en este caso es:

$$V(\underline{x}) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

mientras que la matriz de varianzas y covarianzas de los datos simulados es:

$$V_{sim}(\underline{x}) = \begin{bmatrix} 1.003 & -0.014 \\ -0.014 & 1.017 \end{bmatrix}$$

Por su parte, la distribución Pearson Tipo VII tiene una media poblacional dada por:

$$E(\underline{x}) = (0, 0)$$

y la media de los datos simulados es:

$$\bar{\underline{x}} = (0.006, 0.014)$$

La matriz de varianzas y covarianzas poblacional para este caso es:

$$V(\underline{x}) = \begin{bmatrix} 0.937 & 0.000 \\ 0.000 & 0.937 \end{bmatrix}$$

mientras que la matriz de varianzas y covarianzas de los datos simulados es:

$$V_{sim}(\underline{x}) = \begin{bmatrix} 0.958 & -0.002 \\ -0.002 & 0.949 \end{bmatrix}$$

Estos resultados corroboran las similitudes en las gráficas y la viabilidad de usar este método de simulación.

Por último se presentan los resultados de una simulación para una distribución Pearson Tipo VII en dimensión 4 con parámetros $\gamma = 14$ y $m = 10$, media poblacional:

$$E(\underline{x}) = (1, 0, -1, 0)$$

y matriz de varianzas y covarianzas poblacional

$$V(\underline{x}) = \begin{bmatrix} 1 & 0.5 & 0.25 & 0.125 \\ 0.5 & 1 & 0.5 & 0.25 \\ 0.25 & 0.5 & 1 & 0.5 \\ 0.125 & 0.25 & 0.5 & 1 \end{bmatrix}$$

En este caso la media de los datos simulados resulta ser:

$$\bar{\underline{x}} = (1.016, 0.007, -1.009, -0.006)$$

y la matriz de varianzas y covarianzas:

$$V_{sim}(\underline{x}) = \begin{bmatrix} 0.992 & 0.5 & 0.245 & 0.123 \\ 0.5 & 1.015 & 0.507 & 0.25 \\ 0.245 & 0.507 & 1.004 & 0.504 \\ 0.123 & 0.25 & 0.504 & 1 \end{bmatrix}$$

Es notoria la similitud entre los resultados poblacionales y los obtenidos por la simulación, por lo cual se constata que el algoritmo funciona razonablemente bien.

Capítulo IV. Análisis de Componentes Principales Comunes en Vectores Dependientes Bajo el Supuesto de Distribución Elíptica

Hasta ahora, los tres capítulos previos han presentado un conjunto de resultados que permiten establecer el contexto en el que se enmarca el presente proyecto de investigación. Si bien dichos capítulos son un compendio estructurado de resultados relacionados con el análisis de muestras complejas, el Análisis de Componentes Principales Comunes en Vectores Dependientes y las Distribuciones Elípticas, que no se encuentra reunido en ningún texto hasta ahora escrito, ello no conlleva una propuesta que pueda valorarse como un proyecto de investigación original.

En los siguientes dos capítulos, se mostrará que los resultados presentados hasta ahora, pueden ser utilizados para plantear un modelo que, partiendo del Análisis de Componentes Principales Comunes en Vectores Dependientes descrito en el *Capítulo II*, pueda generalizarse en dos sentidos. El primero, incluido en el presente capítulo, se relaciona con la posibilidad de partir de un supuesto distribucional alternativo al de Normalidad Multivariada, por medio del uso de las distribuciones elípticas presentadas en el *Capítulo III*, destacando que dichas distribuciones conforman una familia más general que la Normal Multivariada, implicando con ello que será más probable que un conjunto de datos pueda ser modelado por medio de alguna de dichas distribuciones elípticas. El segundo sentido de generalización se presenta en el próximo capítulo, buscando generar un modelo que pueda aplicarse cuando se dispone de muestras de diseño complejo.

Estos resultados sí conllevan una propuesta de carácter original, ya que todas las proposiciones y teoremas presentados en los siguientes dos capítulos, son de autoría propia, y hasta donde se tiene conocimiento, no han sido publicados en ningún texto, considerando que se ha hecho una exhaustiva búsqueda al respecto tanto en los textos publicados hasta ahora de manera escrita en los libros y revistas de mayor tradición en el área de análisis correspondiente, como en internet. Adicionalmente, los resultados que se exponen responden a necesidades reales que se presentan al implementar de manera práctica un método como el Análisis de Componentes Principales Comunes en Vectores Dependientes, puesto que existe por un lado, una necesidad de aplicar modelos con validez en los supuestos que permita asegurar que los resultados obtenidos se apegan a la realidad, y por otro, un uso creciente de las encuestas de diseño complejo debido a la optimización de recursos que implican dichos diseños.

En este capítulo se obtienen los estimadores máximo verosímiles bajo el supuesto de distribución elíptica para los parámetros Λ_{ij} y β del modelo del Análisis de Componentes Principales Comunes en Vectores Dependientes, así como las distribuciones asintóticas de dichos estimadores, haciendo particular énfasis en comparar los resultados obtenidos, con los que parten del supuesto de Normalidad Multivariada, con la finalidad de establecer su utilidad.

Se continúa con la estrategia de presentar los resultados base de las proposiciones y teoremas, incluyendo la fuente en la cual se encuentran demostrados, o en su defecto, probándolos cuando no se haya encontrado un texto conteniendo dichas demostraciones.

IV.1 Estimación Puntual de los Parámetros

Una forma de representar la información que puede estudiar el Análisis de Componentes Principales Comunes en Vectores Dependientes consiste en construir un vector conteniendo los datos de las variables de el i -ésimo individuo de cada conjunto de manera concatenada.

Definición IV.1. Sea $x_i = (x_i^{11}, \dots, x_i^{1p}, x_i^{21}, \dots, x_i^{2p}, \dots, x_i^{k1}, \dots, x_i^{kp})$ tal que x_i^r es el valor de la variable r del individuo i perteneciente al conjunto l .

Por el **Teorema III.3** se tiene que la verosimilitud de la matriz $X = (x_1 \ x_2 \ \dots \ x_n)'$, para una muestra aleatoria x_1, x_2, \dots, x_n , está dada por:

$$L(\underline{\mu}, \Sigma) = |\Sigma|^{-\frac{n}{2}} g^*(tr \Sigma^{-1} G) \quad (IV.1)$$

En [Fang y Anderson, 1990a] se demuestra que este resultado es equivalente a:

$$L(\underline{\mu}, \Sigma) = |\Sigma|^{-\frac{n}{2}} g^*(tr \Sigma^{-1} W + n(\bar{X} - \underline{\mu})' \Sigma^{-1} (\bar{X} - \underline{\mu})) \quad (IV.2)$$

Recordando que $\Psi = \frac{E(R^2)}{nkp} \Sigma$ se tiene entonces que $\Sigma = \varrho \Psi$ con $\varrho = \frac{nkp}{E(R^2)}$.

Dado que, bajo el modelo previamente planteado se tiene que $\Psi_{ij} = \beta \Lambda_{ij} \beta'$, entonces $\Sigma = [\Sigma_{ij}]_{i,j=1,\dots,k}$ se puede expresar como $\Sigma_{ij} = [\varrho \beta \Lambda_{ij} \beta']_{i,j=1,\dots,k}$ de modo que la verosimilitud original puede ser reparametrizada en términos de $\underline{\mu}$, ϱ , β y Λ_{ij} . Si se define a Λ como la matriz por bloques conformada como $\Lambda = [\Lambda_{ij}]_{i,j=1,\dots,k}$ entonces, bajo el modelo, se puede escribir a Σ como $\Sigma = \varrho [I_k \otimes \beta] \Lambda [I_k \otimes \beta]'$, donde I_k es la matriz identidad de dimensión k .

Lema IV.1. Sea la matriz Σ tal que $\Sigma = \varrho[I_k \otimes \beta]\Lambda[I_k \otimes \beta]'$ entonces:

$$|\Sigma| = \varrho^{kp} |\Lambda| \quad (IV.3)$$

Demostración:

$$\begin{aligned} |\Sigma| &= |\varrho[I_k \otimes \beta]\Lambda[I_k \otimes \beta]'| \\ &= \varrho^{kp} |[I_k \otimes \beta]\Lambda[I_k \otimes \beta]'| \\ &= \varrho^{kp} |[I_k \otimes \beta]| |\Lambda| |[I_k \otimes \beta]'| \\ &= \varrho^{kp} |[I_k \otimes \beta]|^2 |\Lambda| \\ &= \varrho^{kp} |I_k|^{2p} |\beta|^{2k} |\Lambda| \\ &= \varrho^{kp} 1^{2p} (\pm 1)^{2k} |\Lambda| \\ &= \varrho^{kp} |\Lambda| \end{aligned}$$

QED

Lema IV.2. Sea la matriz Σ tal que $\Sigma = \varrho[I_k \otimes \beta]\Lambda[I_k \otimes \beta]'$ entonces:

$$tr(\Sigma^{-1}W) = \varrho^{-1} tr(\Lambda^{-1}D) \quad (IV.4)$$

donde $D = [I_k \otimes \beta]'W[I_k \otimes \beta]$.

Demostración:

$$\begin{aligned} tr(\Sigma^{-1}W) &= tr\left(\varrho^{-1}([I_k \otimes \beta]\Lambda[I_k \otimes \beta]')^{-1}W\right) \\ &= \varrho^{-1} tr\left([I_k \otimes \beta]'^{-1}\Lambda^{-1}[I_k \otimes \beta]^{-1}W\right) \\ &= \varrho^{-1} tr\left([I_k^{-1} \otimes \beta^{-1}]'\Lambda^{-1}[I_k^{-1} \otimes \beta^{-1}]W\right) \\ &= \varrho^{-1} tr\left([I_k' \otimes \beta']'\Lambda^{-1}[I_k' \otimes \beta']W\right) \\ &= \varrho^{-1} tr\left(\Lambda^{-1}[I_k' \otimes \beta']W[I_k' \otimes \beta']'\right) \\ &= \varrho^{-1} tr\left(\Lambda^{-1}[I_k \otimes \beta]'W[I_k \otimes \beta]\right) \end{aligned}$$

QED

Por los **Lemas IV.1 y IV.2**, se puede reescribir la verosimilitud como:

$$L(\underline{\mu}, \beta, \Lambda, \varrho) = \varrho^{-\frac{nkp}{2}} |\Lambda|^{-\frac{n}{2}} g^*(\varrho^{-1} tr(\Lambda^{-1}D) + n(\bar{X} - \underline{\mu})' \Sigma^{-1}(\bar{X} - \underline{\mu})) \quad (IV.5)$$

La función cuadrática $n(\bar{X} - \underline{\mu})' \Sigma^{-1}(\bar{X} - \underline{\mu})$ es el único término de la verosimilitud que depende de $\underline{\mu}$, y tiene como único punto crítico a $\hat{\underline{\mu}} = \bar{X}$, independientemente del valor de Σ , por tanto maximizar la verosimilitud de la expresión (IV.5) particularmente para los parámetros β y Λ es equivalente a maximizar la verosimilitud:

$$L(\bar{X}, \beta, \Lambda, \varrho) = \varrho^{-\frac{nkp}{2}} |\Lambda|^{-\frac{n}{2}} g^*(\varrho^{-1} tr(\Lambda^{-1}D)) \quad (IV.6)$$

Para encontrar los estimadores de máxima verosimilitud de Λ y β se hace uso de los siguientes teoremas de cálculo matricial.

Lema IV.3. Sea Z una matriz simétrica con entradas funcionalmente independientes (excepto por $Z_{ij} = Z_{ji}$), y A una matriz real, entonces:

$$\frac{\partial}{\partial Z} tr(AZ) = A + A' - diag(A) \quad (IV.7)$$

$$\frac{\partial}{\partial Z} \log |Z| = 2Z^{-1} - diag(Z^{-1}) \quad (IV.8)$$

Demostración: se encuentra en [Henderson y Searle, 1979, pag. 76].

Lema IV.4. Sean C_j matrices diagonales $C_j = diag(c_{j1}, \dots, c_{jp})$, y sean A_j matrices de dimensión $p \times p$, $j = 1, \dots, R$. Sea además $\beta = [\beta_1, \dots, \beta_p]$ una matriz ortogonal de dimensión $p \times p$. Entonces los puntos estacionarios de la función

$$f(\beta) = \sum_{j=1}^R tr(C_j \beta' A_j \beta)$$

son las soluciones de

$$\beta'_m \left[\sum_{j=1}^R (c_{jl} - c_{jm}) (A_j + A'_j) \right] \beta_l = 0 \quad (IV.9)$$

para toda $l, m = 1, \dots, p$ ($l \neq m$)

Demostración: se encuentra en [Neuenschwander y Flury, 2000, pag. 179].

En la **Sección II.4.2** se definió la matriz Λ^* a partir de la transformación $\Lambda^* = I_{p,k} \Lambda'_{p,k}$, siendo $I_{p,k}$ matrices de conmutación (que en particular son ortogonales). A continuación se presentan dos resultados que relacionan operaciones matriciales involucrando a Λ y a Λ^* .

Lema IV.5. Sean Λ y Λ^* tales que $\Lambda^* = I_{p,k} \Lambda'_{p,k}$ entonces:

$$|\Lambda^{*-1}| = |\Lambda^{-1}| \quad (IV.10)$$

Demostración:

Dado que $I_{p,k}$ es ortogonal, entonces es directo probar que $|\Lambda^*| = |\Lambda|$, por la conmutatividad de los operadores inversa y determinante se deduce el resultado.

QED

Lema IV.6. Sean Λ y Λ^* tales que $\Lambda^* = I_{p,k} \Lambda'_{p,k}$ entonces:

$$\text{tr}(\Lambda^{*-1}Z) = \text{tr}(\Lambda^{-1}D) \quad (IV.11)$$

con:

$$Z = I_{p,k} D I'_{p,k} \quad (IV.12)$$

Demostración:

$$\begin{aligned} \text{tr}(\Lambda^{*-1}Z) &= \text{tr}([I_{p,k} \Lambda'_{p,k}]^{-1} I_{p,k} D I'_{p,k}) \\ &= \text{tr}(I'^{-1}_{p,k} \Lambda^{-1} I^{-1}_{p,k} I_{p,k} D I'_{p,k}) \\ &= \text{tr}(\Lambda^{-1} D I'_{p,k} I'^{-1}_{p,k}) \\ &= \text{tr}(\Lambda^{-1} D) \end{aligned}$$

QED

Lema IV.7. Para Λ^* se tiene que:

$$\log |\Lambda^{*-1}| = \sum_{h=1}^p \log |\Lambda_h^{*-1}| \quad (IV.13)$$

y

$$\text{tr}(\Lambda^{*-1}Z) = \sum_{h=1}^p \text{tr}(\Lambda_h^{*-1}Z_{hh}) \quad (IV.14)$$

con $Z = [Z_{ij}]_{i,j=1,\dots,p}$.

Demostración: se deriva directamente de las propiedades de las matrices de forma diagonal por bloques.

Teorema IV.1. Sea $\underline{x}_1, \dots, \underline{x}_n$ una muestra aleatoria de vectores de acuerdo con la definición IV.1, teniendo distribución elíptica $\underline{x}_i \sim EC_p(\underline{\mu}, \Sigma, g)$, donde la función $g^*(t)$ que la caracteriza sea decreciente y diferenciable, entonces los estimadores máximo verosímiles de Λ_{ij} y β para el modelo $\Psi_{ij} = \beta \Lambda_{ij} \beta'$ con Ψ_{ij} matrices de varianzas y covarianzas, satisfacen el siguiente sistema de ecuaciones, sujetos a la restricción de ortogonalidad de β :

$$\lambda_{ij,h} = u \beta'_h W_{ij} \beta_h, \quad i, j = 1, \dots, k, \quad h = 1, \dots, p \quad (IV.15)$$

$$\beta'_h \left[\sum_{i=1}^k \sum_{j=1}^k (\lambda^{ij,l} - \lambda^{ij,h}) (W_{ij}) \right] \beta_l = 0 \quad (IV.16)$$

donde

$$u = -\frac{2}{n} \frac{\frac{E(R^{*2})}{nk\rho} g^* \left(\frac{E(R^{*2})}{u} \right)}{g^* \left(\frac{E(R^{*2})}{u} \right)} \quad (IV.17)$$

Demostración:

Reescribiendo la verosimilitud de la ecuación (IV.6) en términos de Λ^{-1} se tiene:

$$L(\bar{X}, \beta, \Lambda^{-1}, \varrho) = \varrho^{-\frac{nk\rho}{2}} |\Lambda^{-1}|^{\frac{n}{2}} g^*(\varrho^{-1} \text{tr}(\Lambda^{-1}D))$$

La función log-verosimilitud es entonces:

$$\mathcal{L} = \log L(\bar{X}, \beta, \Lambda^{-1}, \varrho) = C + \frac{n}{2} \log(|\Lambda^{-1}|) + \log[g^*(\varrho^{-1} \text{tr}(\Lambda^{-1} D))]$$

Si se utiliza a la matriz Λ^* , entonces por los **Lemas IV.5 y IV.6** se tiene que:

$$\mathcal{L}^* = \log L^*(\bar{X}, \beta, \Lambda^{*-1}, \varrho) = C + \frac{n}{2} \log(|\Lambda^{*-1}|) + \log[g^*(\varrho^{-1} \text{tr}(\Lambda^{*-1} D))]$$

el **Lema IV.7** implica que:

$$\mathcal{L}^* = \log L^*(\bar{X}, \beta, \Lambda^{*-1}, \varrho) = C + \frac{n}{2} \sum_{h=1}^p \log |\Lambda_h^{*-1}| + \log[g^*(\varrho^{-1} \sum_{h=1}^p \text{tr}(\Lambda_h^{*-1} Z_{hh}))]$$

donde cada Λ_h^* es una matriz simétrica con elementos funcionalmente independientes excepto por $(\Lambda_h^*)_{ij} = (\Lambda_h^*)_{ji}$. Por el **Lema IV.3** se tiene:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \Lambda_h^{*-1}} &= \frac{n}{2} [2\Lambda_h^* - \text{diag}(\Lambda_h^*)] + \frac{g^{*\prime}(\varrho^{-1} \text{tr}(\Lambda_h^{*-1} Z_{hh})) (\varrho^{-1} (Z_{hh} + Z'_{hh} - \text{diag}(Z_{hh})))}{g^*(\varrho^{-1} \text{tr}(\Lambda_h^{*-1} Z_{hh}))} \\ &= \frac{n}{2} \left[2\Lambda_h^* - \text{diag}(\Lambda_h^*) + \frac{2\varrho^{-1} g^{*\prime}(\varrho^{-1} \text{tr}(\Lambda_h^{*-1} Z_{hh}))}{ng^*(\varrho^{-1} \text{tr}(\Lambda_h^{*-1} Z_{hh}))} (2Z_{hh} - \text{diag}(Z_{hh})) \right] \end{aligned}$$

Igualando a cero se obtiene:

$$2\widehat{\Lambda}_h^* - 2u_h Z_{hh} + \text{diag}(u_h Z_{hh}) - \text{diag}(\widehat{\Lambda}_h^*) = 0$$

para

$$u_h = \frac{-2\varrho^{-1} g^{*\prime}(\varrho^{-1} \text{tr}(\Lambda_h^{*-1} Z_{hh}))}{ng^*(\varrho^{-1} \text{tr}(\Lambda_h^{*-1} Z_{hh}))} \quad h = 1, \dots, p$$

de donde

$$\widehat{\Lambda}_h^* = u_h Z_{hh}$$

lo que implica que:

$$\lambda_{ij,h} = u_h \beta'_h W_{ij} \beta_h, \quad i, j = 1, \dots, k, \quad h = 1, \dots, p$$

Para encontrar el valor de u_h se puede sustituir el valor de $\widehat{\Lambda}_h^* = u_h Z_{hh}$ en la ecuación anterior, con lo cual se obtiene la ecuación:

$$u_h = \frac{-2\rho^{-1}g^{*\prime}(\rho^{-1}u_h^{-1}k\rho)}{ng^*(\rho^{-1}u_h^{-1}k\rho)}$$

Sustituyendo el valor de ρ se obtiene:

$$u_h = -\frac{2}{n} \frac{\frac{E(R^{*2})}{nkp} g^{*\prime}\left(\frac{E(R^{*2})}{nu_h}\right)}{g^*\left(\frac{E(R^{*2})}{nu_h}\right)} \quad \forall h$$

Debido a que el subíndice h solo está asociado al término u_h , se puede definir entonces a $u = u_h$ de modo que:

$$\lambda_{ij,h} = u\beta'_h W_{ij}\beta_h, \quad i, j = 1, \dots, k, \quad h = 1, \dots, p$$

Para estimar β se regresa a la verosimilitud original, y se usa el hecho de que la función $g^*(h(x))$ sea decreciente, por tanto hallar el óptimo para x en $\log(g^*(h(x)))$ es equivalente a hallar el óptimo en $h(x)$. En ese caso basta con hallar el óptimo de la función $n\rho^{-1}tr(\Lambda^{-1}D)$, dentro de la cual $n\rho^{-1}$ es una constante con respecto a β por lo cual es suficiente con encontrar el óptimo de $tr(\Lambda^{-1}D)$ como función de β . Una forma alternativa de escribir dicha función está dada por:

$$tr(\Lambda^{-1}D) = \sum_{i=1}^k \sum_{j=1}^k tr(\Lambda^{ij}\beta'W_{ij}\beta)$$

donde $\Lambda^{-1} = [\Lambda^{ij}]_{i,j=1,\dots,k}$.

Como Λ_{ij} son diagonales en Λ , entonces Λ^{ij} serán diagonales en Λ^{-1} . Para encontrar el estimador de β se usa el **Lema IV.4** del cual se desprende que los puntos estacionarios de $tr(\Lambda^{-1}D)$ son las soluciones de:

$$\beta'_h \left[\sum_{i=1}^k \sum_{j=1}^k (\lambda^{ij,l} - \lambda^{ij,m})(W_{ij}) \right] \beta_l = 0$$

donde $\lambda^{ij,l}$ son los elementos diagonales de Λ^{ij} .

QED

El sistema de ecuaciones definido por el **Teorema IV.1** puede resolverse utilizando el algoritmo FG+ de Flury y Neuenschwander con una modificación que considere la presencia del factor u .

A continuación se ejemplifica la obtención del factor u para el caso de la distribución Pearson Tipo VII.

Proposición IV.1. Sea $\underline{x}_1, \dots, \underline{x}_n$ una muestra aleatoria de vectores independientes con idéntica distribución Pearson Tipo VII $\underline{x}_i \sim PP_p(\underline{\mu}, \gamma, m, \Sigma, g)$, entonces el factor u necesario para determinar los estimadores máximo verosímiles de Λ_{ij} y β para el modelo $\Psi_{ij} = \beta \Lambda_{ij} \beta'$ con Ψ_{ij} matrices de varianzas y covarianzas, es:

$$u = \left(\frac{2m - kp}{2m - kp - 2} \right) \frac{1}{n} \quad (IV.18)$$

Demostración:

Dado que $E(R^{*2}) = nE(R^2) = \frac{\gamma nkp}{2m - kp - 2}$ y que $\frac{g^{*'}(t)}{g^*(t)} = \frac{-m^*}{\gamma + t}$, con $m^* = \frac{2m + nkp - kp}{2}$ entonces el valor de u está dado por:

$$u = -\frac{2}{n} \frac{\frac{\gamma nkp}{2m - kp - 2}}{nkp} \left(\frac{-m^*}{\gamma + \left(\frac{\frac{\gamma nkp}{2m - kp - 2}}{nu} \right)} \right)$$

de donde:

$$\begin{aligned} u &= \frac{\frac{2m^* \gamma}{2m - kp - 2} - \frac{\gamma nkp}{2m - kp - 2}}{n\gamma} \\ &= \left(\frac{2m^* - nkp}{2m - kp - 2} \right) \frac{1}{n} \\ &= \left(\frac{2m + nkp - kp - nkp}{2m - kp - 2} \right) \frac{1}{n} \\ &= \left(\frac{2m - kp}{2m - kp - 2} \right) \frac{1}{n} \end{aligned}$$

QED

Proposición IV.2. La prueba de cociente de verosimilitudes generalizado para definir estadísticamente la validez del modelo de Análisis de Componentes Principales Comunes en Vectores Dependientes bajo el supuesto de distribución elíptica tiene como estadístico de prueba a:

$$\chi^2_{ACPCD} = n \left[\log \left(\frac{|pdiag(\hat{\Lambda})|}{|\hat{\Lambda}|} \right) \right] \quad (IV.19)$$

donde, bajo la hipótesis de existencia de las Componentes Principales Comunes en Vectores Dependientes, χ^2_{ACPCD} se distribuye asintóticamente como una χ^2 con $(k^2 - 1) \left(\frac{p(p-1)}{2} \right)$ con

$$pdiag(\hat{\Lambda}) = \begin{bmatrix} diag(\hat{\Lambda}_{11}) & \dots & diag(\hat{\Lambda}_{1k}) \\ \vdots & \ddots & \vdots \\ diag(\hat{\Lambda}_{k1}) & \dots & diag(\hat{\Lambda}_{kk}) \end{bmatrix}$$

Demostración:

La verosimilitud es:

$$L(\underline{\mu}, \underline{\Sigma}) = |\underline{\Sigma}|^{-\frac{n}{2}} g^* \left(tr(\underline{\Sigma}^{-1}W + n(\underline{\bar{x}} - \underline{\mu})' \underline{\Sigma}^{-1}(\underline{\bar{x}} - \underline{\mu})) \right)$$

Evaluada en $\underline{\bar{x}}$ y $\hat{\underline{\Sigma}}$ es:

$$L(\underline{\bar{x}}, \hat{\underline{\Sigma}}) = |\hat{\underline{\Sigma}}|^{-\frac{n}{2}} g^* \left(tr(\hat{\underline{\Sigma}}^{-1}W) \right)$$

Bajo la hipótesis de que el modelo del Análisis de Componentes Principales Comunes en Vectores Dependientes se cumple, se tiene:

$$\hat{\underline{\Sigma}} = \varrho \hat{\underline{\Psi}} = \varrho [I_k \otimes \hat{\beta}] \hat{F}_{ACPCD} [I_k \otimes \hat{\beta}]'$$

para F_{ACPCD} diagonal.

Bajo ninguna restricción se tiene:

$$\hat{\underline{\Sigma}} = \varrho \hat{\underline{\Psi}} = \varrho [I_k \otimes \hat{\beta}] \hat{F} [I_k \otimes \hat{\beta}]'$$

para F no necesariamente diagonal.

Sean $\hat{F}_{ACPCD} = pdiag(\hat{\Lambda})$ y $\hat{F} = \hat{\Lambda}$. Se sabe que $\hat{\underline{\Sigma}} = u \varrho W$ con u definida en el **Teorema IV.1**, por lo tanto:

$$\begin{aligned}
L_{ACPCD}(\bar{\mathbf{x}}, \hat{\boldsymbol{\Sigma}}) &= |\mathbf{u}\boldsymbol{\varrho}\mathbf{W}|^{-\frac{n}{2}} g^* \left(\text{tr}(\mathbf{u}^{-1}\boldsymbol{\varrho}^{-1}\mathbf{W}^{-1}\mathbf{W}) \right) \\
&= |\hat{\boldsymbol{\Sigma}}|^{-\frac{n}{2}} g^* \left(\text{tr}(\mathbf{u}^{-1}\boldsymbol{\varrho}^{-1}\mathbf{I}_{kp}) \right) \\
&= |\boldsymbol{\varrho}| |\mathbf{I}_k \otimes \hat{\boldsymbol{\beta}}| p \text{diag}(\hat{\boldsymbol{\Lambda}}) |\mathbf{I}_k \otimes \hat{\boldsymbol{\beta}}'|^{-\frac{n}{2}} g^* \left(\text{tr}(\mathbf{u}^{-1}\boldsymbol{\varrho}^{-1}\mathbf{I}_{kp}) \right) \\
&= |\boldsymbol{\varrho}|^{-\frac{n}{2}} |\mathbf{I}_k \otimes \hat{\boldsymbol{\beta}}|^{-\frac{n}{2}} |p \text{diag}(\hat{\boldsymbol{\Lambda}})|^{-\frac{n}{2}} |\mathbf{I}'_k \otimes \hat{\boldsymbol{\beta}}'|^{-\frac{n}{2}} g^* \left(\text{tr}(\mathbf{u}^{-1}\boldsymbol{\varrho}^{-1}\mathbf{I}_{kp}) \right) \\
&= |\boldsymbol{\varrho}|^{-\frac{n}{2}} |p \text{diag}(\hat{\boldsymbol{\Lambda}})|^{-\frac{n}{2}} g^* \left(\text{tr}(\mathbf{u}^{-1}\boldsymbol{\varrho}^{-1}\mathbf{I}_{kp}) \right)
\end{aligned}$$

y

$$\begin{aligned}
L(\bar{\mathbf{x}}, \hat{\boldsymbol{\Sigma}}) &= |\mathbf{u}\boldsymbol{\varrho}\mathbf{W}|^{-\frac{n}{2}} g^* \left(\text{tr}(\mathbf{u}^{-1}\boldsymbol{\varrho}^{-1}\mathbf{W}^{-1}\mathbf{W}) \right) \\
&= |\hat{\boldsymbol{\Sigma}}|^{-\frac{n}{2}} g^* \left(\text{tr}(\mathbf{u}^{-1}\boldsymbol{\varrho}^{-1}\mathbf{I}_{kp}) \right) \\
&= |\boldsymbol{\varrho}| |\mathbf{I}_k \otimes \hat{\boldsymbol{\beta}}| \hat{\boldsymbol{\Lambda}} |\mathbf{I}_k \otimes \hat{\boldsymbol{\beta}}'|^{-\frac{n}{2}} g^* \left(\text{tr}(\mathbf{u}^{-1}\boldsymbol{\varrho}^{-1}\mathbf{I}_{kp}) \right) \\
&= |\boldsymbol{\varrho}|^{-\frac{n}{2}} |\mathbf{I}_k \otimes \hat{\boldsymbol{\beta}}|^{-\frac{n}{2}} |\hat{\boldsymbol{\Lambda}}|^{-\frac{n}{2}} |\mathbf{I}'_k \otimes \hat{\boldsymbol{\beta}}'|^{-\frac{n}{2}} g^* \left(\text{tr}(\mathbf{u}^{-1}\boldsymbol{\varrho}^{-1}\mathbf{I}_{kp}) \right) \\
&= |\boldsymbol{\varrho}|^{-\frac{n}{2}} |\hat{\boldsymbol{\Lambda}}|^{-\frac{n}{2}} g^* \left(\text{tr}(\mathbf{u}^{-1}\boldsymbol{\varrho}^{-1}\mathbf{I}_{kp}) \right)
\end{aligned}$$

de donde el cociente de verosimilitudes es:

$$\begin{aligned}
\frac{L_{ACPCD}(\bar{\mathbf{x}}, \hat{\boldsymbol{\Sigma}})}{L(\bar{\mathbf{x}}, \hat{\boldsymbol{\Sigma}})} &= \frac{|\boldsymbol{\varrho}|^{-\frac{n}{2}} |p \text{diag}(\hat{\boldsymbol{\Lambda}})|^{-\frac{n}{2}} g^* \left(\text{tr}(\mathbf{u}^{-1}\boldsymbol{\varrho}^{-1}\mathbf{I}_{kp}) \right)}{|\boldsymbol{\varrho}|^{-\frac{n}{2}} |\hat{\boldsymbol{\Lambda}}|^{-\frac{n}{2}} g^* \left(\text{tr}(\mathbf{u}^{-1}\boldsymbol{\varrho}^{-1}\mathbf{I}_{kp}) \right)} \\
&= \left(\frac{|p \text{diag}(\hat{\boldsymbol{\Lambda}})|}{|\hat{\boldsymbol{\Lambda}}|} \right)^{-\frac{n}{2}}
\end{aligned}$$

Por lo tanto:

$$-2 \log \left(\frac{L_{ACPCD}(\bar{\mathbf{x}}, \hat{\boldsymbol{\Sigma}})}{L(\bar{\mathbf{x}}, \hat{\boldsymbol{\Sigma}})} \right) = n \left[\log \left(\frac{|p \text{diag}(\hat{\boldsymbol{\Lambda}})|}{|\hat{\boldsymbol{\Lambda}}|} \right) \right]$$

QED

IV.2 Distribución Asintótica de los Parámetros

Los estimadores $\hat{\Lambda}$ y $\hat{\beta}$ se obtienen a partir de expresiones que no son explícitas a partir de un sistema de ecuaciones, por lo que no es posible encontrar sus distribuciones exactas. Por ello se estudiarán sus distribuciones asintóticas.

Lema IV.8. Sean G_m , H_m y $I_{m,m}$ las matrices de duplicación, eliminación y permutación respectivamente, en el contexto definido por Henderson y Searle (1979), donde para $A_{m \times m}$ simétrica se tiene que $vec(A) = G_m vech(A)$, $vech(A) = H_m vec(A)$ y $vec(A) = (I_{m,m}) vec(A')$, y donde $(I_{m,m})(I_{m,m}) = I_{m^2}$. Sea además $D_m := diag(vec(I_m))$, entonces:

$$\begin{aligned}
 i) & H_m(2I_{m^2} - D_m) = G'_m \\
 ii) & H_m(I_{m^2} + I_{m,m} - D_m) = G'_m \\
 iii) & G'_m(A \otimes A)(I_{m^2} + I_{m,m}) = 2G'_m(A \otimes A) \\
 iv) & H_m(A \otimes A)(I_{m^2} + I_{m,m}) = 2H_m(A \otimes A)
 \end{aligned} \tag{IV.20}$$

Demostración: se encuentra en [Neuenschwander y Flury, 2000, pag. 180].

Lema IV.9. Sea $A_{m \times m}$ simétrica con variables funcionalmente independientes (excepto por $A_{ij} = A_{ji}$), entonces:

$$\begin{aligned}
 i) & \frac{\partial vec(A^{-1})}{\partial (vec(A))'} = -(A^{-1} \otimes A^{-1})(I_{m^2} + I_{m,m} - D_m) \\
 ii) & \frac{\partial vech(A^{-1})}{\partial (vech(A))'} = -H_m(A^{-1} \otimes A^{-1})G_m
 \end{aligned} \tag{IV.21}$$

Demostración: se encuentra en [Neuenschwander y Flury, 2000, pag. 181].

Definición IV.2. Las condiciones de regularidad en el caso multivariado son:

(R1) Para cada $\theta \in \Theta$ las siguientes derivadas existen para toda \underline{x} :

$$\begin{aligned}
 & \frac{\partial}{\partial \theta_i} \log f(\underline{x} : \theta) \quad \forall i \\
 & \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log f(\underline{x} : \theta) \quad \forall i, j \\
 & \frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \theta_k} \log f(\underline{x} : \theta) \quad \forall i, j, k
 \end{aligned}$$

(R2) Existen funciones $g(\underline{x})$, $h(\underline{x})$ y $H(\underline{x})$ Para cada $\underline{\theta}_0 \in \Theta$ (posiblemente dependiendo de $\underline{\theta}_0$) tales que, para $\underline{\theta}$ en una vecindad $N(\underline{\theta}_0)$, las siguientes desigualdades se cumplen para toda \underline{x} :

$$\left| \frac{\partial}{\partial \theta_i} f(\underline{x} : \underline{\theta}) \right| \leq g(\underline{x}) \quad \forall i$$

$$\left| \frac{\partial^2}{\partial \theta_i \partial \theta_j} f(\underline{x} : \underline{\theta}) \right| \leq h(\underline{x}) \quad \forall i, j$$

$$\left| \frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \theta_k} f(\underline{x} : \underline{\theta}) \right| \leq H(\underline{x}) \quad \forall i, j, k$$

$$\int \dots \int g(\underline{x}) dx_1 \dots dx_n < \infty$$

$$\int \dots \int h(\underline{x}) dx_1 \dots dx_n < \infty$$

$$E_{\underline{\theta}} \{H(\underline{x})\} < \infty$$

(R3) Para cada $\underline{\theta} \in \Theta$ la siguiente desigualdad se cumple para toda \underline{x} :

$$0 < -E_{\underline{\theta}} \left\{ \left(\frac{\partial}{\partial \theta_i} \log f(\underline{x} : \underline{\theta}) \right) \left(\frac{\partial}{\partial \theta_j} \log f(\underline{x} : \underline{\theta}) \right) \right\} < \infty \quad \forall i, j$$

Una vez definidos estos resultados se plantea hacer uso de la propiedad de los estimadores máximo verosímiles que establece que, bajo condiciones de regularidad, la distribución asintótica de un estimador $\hat{\underline{\theta}}$ para un vector de parámetros $\underline{\theta}$ de dimensión p está dada por: $\sqrt{n}(\hat{\underline{\theta}} - \underline{\theta}) \sim N_p(\mathbf{0}, nI_n^{-1}(\underline{\theta}))$ donde $I_n(\underline{\theta})$ es la información esperada de Fisher para $\underline{\theta}$.

Para ello se definen los siguientes vectores: $\underline{\theta}'_1 = [\text{vech}(\Lambda_1^*), \dots, \text{vech}(\Lambda_p^*)]$, $\underline{\theta}'_2 = \underline{\beta}^*$ donde $\underline{\beta}^*$ es el vector integrado por $\frac{p(p-1)}{2}$ elementos funcionalmente independientes de β , y donde $\underline{\phi}' = [\text{vech}(\Lambda_1^{*-1}), \dots, \text{vech}(\Lambda_p^{*-1})]$. El objetivo es encontrar $I_n(\underline{\theta}_1)$ y $I_n(\underline{\theta}_2)$. Para encontrar $I_n(\underline{\theta}_1)$ se encuentra primero $I_n(\underline{\phi})$ y se hace uso de la relación:

$$I_n(\underline{\theta}_1) = \left(\frac{\partial \underline{\phi}'}{\partial \underline{\theta}_1} \right) I_n(\underline{\phi}) \left(\frac{\partial \underline{\phi}}{\partial \underline{\theta}'_1} \right)$$

Por la forma que tiene el vector $\underline{\phi}$ se puede particionar a $I_n(\underline{\phi})$ como:

$$I_n(\underline{\phi}) = \begin{bmatrix} I_n(\text{vech}(\Lambda_1^{*-1})) & \dots & I_n(\text{vech}(\Lambda_1^{*-1}), \text{vech}(\Lambda_p^{*-1})) \\ \vdots & \ddots & \vdots \\ I_n(\text{vech}(\Lambda_p^{*-1}), \text{vech}(\Lambda_1^{*-1})) & \dots & I_n(\text{vech}(\Lambda_p^{*-1})) \end{bmatrix}$$

Teorema IV.2. Las matrices de información esperada de Fisher correspondientes a $\text{vech}(\Lambda_h^{*-1})$ con $h = 1, \dots, p$; y correspondientes a $(\text{vech}(\Lambda_l^{*-1}), \text{vech}(\Lambda_m^{*-1}))$ con $m, l = 1, \dots, p$ y $m \neq l$ para una distribución elíptica donde $Z = I_{(p,k)}[I_k \otimes \beta]'W[I_k \otimes \beta]I'_{(p,k)}$, son respectivamente:

$$I_n(\text{vech}(\Lambda_h^{*-1})) = \frac{n}{2}G'_k(\Lambda_h^* \otimes \Lambda_h^*)G_k - E(y_1(W)) \quad h = 1, \dots, p \quad (IV.22)$$

$$I_n(\text{vech}(\Lambda_m^{*-1}), \text{vech}(\Lambda_l^{*-1})) = -E(y_2(W)) \quad \begin{matrix} m, l = 1, \dots, p \\ m \neq l \end{matrix} \quad (IV.23)$$

donde

$$y_1(W) = \left[\frac{g^{*''}(\varrho^{-1}(\text{tr}(\Lambda^{*-1}Z)))}{g^*(\varrho^{-1}(\text{tr}(\Lambda^{*-1}Z)))} - \left(\frac{g^{*' }(\varrho^{-1}(\text{tr}(\Lambda^{*-1}Z)))}{g^*(\varrho^{-1}(\text{tr}(\Lambda^{*-1}Z)))} \right)^2 \right] \varrho^{-2}G'_k \text{vec}(Z_{hh})\text{vec}(Z_{hh})'G_k$$

y

$$y_2(W) = \left[\frac{g^{*''}(\varrho^{-1}(\text{tr}(\Lambda^{*-1}Z)))}{g^*(\varrho^{-1}(\text{tr}(\Lambda^{*-1}Z)))} - \left(\frac{g^{*' }(\varrho^{-1}(\text{tr}(\Lambda^{*-1}Z)))}{g^*(\varrho^{-1}(\text{tr}(\Lambda^{*-1}Z)))} \right)^2 \right] \varrho^{-2}G'_k \text{vec}(Z_{mm})\text{vec}(Z_{ll})'G_k$$

Demostración:

Los **Lemas IV.5 y IV.6** permiten redefinir la log verosimilitud obtenida en la **Sección IV.1** como:

$$\mathcal{L}^* = \log L(\bar{X}, \beta, \Lambda^{*-1}, \varrho) = C + \frac{n}{2} \log(|\Lambda^{*-1}|) + \log[g^*(\varrho^{-1} \text{tr}(\Lambda^{*-1}Z))]$$

Utilizando el **Lema IV.7**, la log verosimilitud se reescribe como:

$$\mathcal{L}^* = \log L(\bar{X}, \beta, \Lambda^{*-1}, \varrho) = C + \frac{n}{2} \sum_{j=1}^p \log(|\Lambda_j^{*-1}|) + \log[g^*(\varrho^{-1} \sum_{j=1}^p \text{tr}(\Lambda_j^{*-1} Z_{jj}))]$$

Aplicando el **Lema IV.3** se obtiene:

$$\frac{\partial \mathcal{L}^*}{\partial \Lambda_h^{*-1}} = \frac{n}{2} (2\Lambda_h^* - \text{diag}(\Lambda_h^*)) + \frac{g^{*'}(\varrho^{-1}(\text{tr}(\Lambda^{*-1} Z)))}{g^*(\varrho^{-1}(\text{tr}(\Lambda^{*-1} Z)))} (\varrho^{-1}(2Z_{hh} - \text{diag}(Z_{hh})))$$

Usando el hecho de que $\text{vec}(\text{diag}(\Lambda_h^*)) = D_k \text{vec}(\text{diag}(\Lambda_h^*))$ se tiene:

$$\frac{\partial \mathcal{L}^*}{\partial \text{vec}(\Lambda_h^{*-1})} = \frac{n}{2} (2I_{k^2} - D_k) \text{vec}(\Lambda_h^*) + \frac{g^{*'}(\varrho^{-1}(\text{tr}(\Lambda^{*-1} Z)))}{g^*(\varrho^{-1}(\text{tr}(\Lambda^{*-1} Z)))} (\varrho^{-1}(2I_{k^2} - D_k) \text{vec}(Z_{hh}))$$

Y por tanto

$$\frac{\partial \mathcal{L}^*}{\partial \text{vech}(\Lambda_h^{*-1})} = \frac{n}{2} H_k (2I_{k^2} - D_k) \text{vec}(\Lambda_h^*) + \frac{g^{*'}(\varrho^{-1}(\text{tr}(\Lambda^{*-1} Z)))}{g^*(\varrho^{-1}(\text{tr}(\Lambda^{*-1} Z)))} (\varrho^{-1} H_k (2I_{k^2} - D_k) \text{vec}(Z_{hh}))$$

Aplicando la regla de la cadena y el **Lema IV.9** se tiene que:

$$\frac{\partial^2 \mathcal{L}^*}{\partial \text{vech}(\Lambda_h^{*-1})' \partial \text{vech}(\Lambda_h^{*-1})} = -\frac{n}{2} H_k (2I_{k^2} - D_k) (\Lambda_h^* \otimes \Lambda_h^*) (I_{k^2} + I_{k,k} - D_k) H_k' + y_1(W)$$

donde

$$y_1(W) = \left[\frac{g^{*''}(\varrho^{-1}(\text{tr}(\Lambda^{*-1} Z)))}{g^*(\varrho^{-1}(\text{tr}(\Lambda^{*-1} Z)))} - \left(\frac{g^{*'}(\varrho^{-1}(\text{tr}(\Lambda^{*-1} Z)))}{g^*(\varrho^{-1}(\text{tr}(\Lambda^{*-1} Z)))} \right)^2 \right] \psi$$

para $\psi = \varrho^{-2} [H_k (2I_{k^2} - D_k) \text{vec}(Z_{hh})] [H_k (2I_{k^2} - D_k) \text{vec}(Z_{hh})]'$

El **Lema IV.8** implica que:

$$\frac{\partial^2 \mathcal{L}^*}{\partial \text{vech}(\Lambda_h^{*-1})' \partial \text{vech}(\Lambda_h^{*-1})} = -\frac{n}{2} G_k' (\Lambda_h^* \otimes \Lambda_h^*) G_k + y_1(W)$$

donde

$$y_1(W) = \left[\frac{g^{**}(\varrho^{-1}(\text{tr}(\Lambda^{*-1}Z)))}{g^*(\varrho^{-1}(\text{tr}(\Lambda^{*-1}Z)))} - \left(\frac{g^{**}(\varrho^{-1}(\text{tr}(\Lambda^{*-1}Z)))}{g^*(\varrho^{-1}(\text{tr}(\Lambda^{*-1}Z)))} \right)^2 \right] \varrho^{-2} G'_k \text{vec}(Z_{hh}) \text{vec}(Z_{hh})' G_k$$

De la definición de información esperada de Fisher se sigue la demostración. Para el caso de $I_n(\text{vech}(\Lambda_j^{*-1}), \text{vech}(\Lambda_m^{*-1}))$ con $m, l = 1, \dots, p$ y $m \neq l$, la prueba es análoga, el primer sumando desaparece al obtener la derivada parcial cruzada.

QED

Como se puede apreciar, no se obtiene una expresión cerrada para las esperanzas cruzadas. Es probable que el conocer la función $g^*(t)$ permita obtener dichas expresiones de manera cerrada como en el caso Normal, donde $\left[\frac{g^{**}(t)}{g^*(t)} - \frac{g^{**2}(t)}{g^{*2}(t)} \right] = 0$, implicando que $y_1(W) = y_2(W) = 0$ (ver expresión (II.29)). A continuación se presenta el caso de la distribución Pearson Tipo VII.

Proposición IV.3. Sean x_1, \dots, x_n vectores aleatorios independientes con idéntica distribución Pearson Tipo VII $x_i \sim PP_p(\mu, \gamma, m, \Sigma, g)$, entonces las esperanzas $E(y_1(W))$ y $E(y_2(W))$ que aparecen en el **Teorema V.1** se pueden calcular como:

$$E(y_1(W)) = \frac{(2m - kp)(2m - kp + 2)}{2\gamma^2(2m + nkp - kp + 2)} \varrho^{-2} G'_k E^{**}(\text{vec}(Z_{hh}) \text{vec}(Z_{hh})') G_k \quad (IV.24)$$

y

$$E(y_2(W)) = \frac{(2m - kp)(2m - kp + 2)}{2\gamma^2(2m + nkp - kp + 2)} \varrho^{-2} G'_k E^{**}(\text{vec}(Z_{mm}) \text{vec}(Z_{ll})') G_k \quad (IV.25)$$

donde $E^{**}(\text{vec}(Z_{hh}) \text{vec}(Z_{hh})')$ y $E^{**}(\text{vec}(Z_{mm}) \text{vec}(Z_{ll})')$ se obtienen a partir de la expresión:

$$E^{**}(ZZ') = I_{p,k} [I_k \otimes \beta]' E^{**}(WW') [I_k \otimes \beta] I'_{p,k} \quad (IV.26)$$

siendo $E^{**}(WW')$ la esperanza de WW' calculada para una distribución Pearson Tipo VII de dimensión nkp con parámetros $\gamma^{**} = \gamma$ y $m^{**} = \frac{2m + nkp - kp}{2} + 2$. Dicha esperanza puede calcularse utilizando el resultado del **Teorema III.5**.

Demostración:

Sean:

$$\begin{aligned} u_1(W) &= \varrho^{-2} G'_k \text{vec}(Z_{hh}) \text{vec}(Z_{hh})' G_k \\ u_2(W) &= \varrho^{-2} G'_k \text{vec}(Z_{mm}) \text{vec}(Z_{ll})' G_k \\ \xi &= \text{tr}(\Sigma^{-1}W) \end{aligned}$$

Recordando que $\text{tr}(\Sigma^{-1}W) = \varrho^{-1}(\text{tr}(\Lambda^{*-1}Z))$, y que $\hat{\mu} = \bar{x}$ es un punto crítico de la verosimilitud independientemente de los valores de los otros parámetros y utilizando el **Teorema III.3**, se tiene que:

$$\begin{aligned} E(y_i(W)) &= \int_X u_i(W) |\Sigma|^{-\frac{n}{2}} g^*(\xi) \left[\frac{g^{*\prime}(\xi)}{g^*(\xi)} - \left(\frac{g^{*\prime}(\xi)}{g^*(\xi)} \right)^2 \right] dX \\ &= \int_X u_i(W) |\Sigma|^{-\frac{n}{2}} \left[g^{*\prime\prime}(\xi) - \frac{g^{*\prime 2}(\xi)}{g^*(\xi)} \right] dX \end{aligned}$$

En el caso de la distribución Pearson Tipo VII se tiene:

$$g^*(t) = \frac{\Gamma(m^*) \left(\frac{1}{\gamma}\right)^{\frac{nkp}{2}}}{\pi^{\frac{nkp}{2}} \Gamma(m^* - \frac{nkp}{2})} \left(1 + \frac{t}{\gamma}\right)^{-m^*}$$

con

$$m^* = \frac{2m + nkp - kp}{2}$$

entonces

$$\begin{aligned} g^{*\prime}(\xi) &= \frac{\Gamma(m^*) \left(\frac{1}{\gamma}\right)^{\frac{nkp}{2}}}{\pi^{\frac{nkp}{2}} \Gamma(m^* - \frac{nkp}{2})} \left(\frac{-m^*}{\gamma}\right) \left(1 + \frac{\xi}{\gamma}\right)^{-m^*-1} \\ g^{*\prime\prime}(\xi) &= \frac{\Gamma(m^*) \left(\frac{1}{\gamma}\right)^{\frac{nkp}{2}}}{\pi^{\frac{nkp}{2}} \Gamma(m^* - \frac{nkp}{2})} \left(\frac{(-m^*)(-m^* - 1)}{\gamma^2}\right) \left(1 + \frac{\xi}{\gamma}\right)^{-m^*-2} \end{aligned}$$

De donde:

$$\begin{aligned}
g^{**}(\xi) - \frac{g^{**2}(\xi)}{g^*(\xi)} &= \frac{\Gamma(m^*) \left(\frac{1}{\gamma}\right)^{\frac{nkp}{2}}}{\pi^{\frac{nkp}{2}} \Gamma(m^* - \frac{nkp}{2})} \left(\frac{m^*}{\gamma^2}\right) \left(1 + \frac{\xi}{\gamma}\right)^{-m^*-2} \\
&= \frac{\left(m^* - \frac{nkp}{2}\right) \left(m^* + 1 - \frac{nkp}{2}\right)}{(m^* + 1)\gamma^2} \left[\frac{\Gamma(m^* + 2) \left(\frac{1}{\gamma}\right)^{\frac{nkp}{2}}}{\pi^{\frac{nkp}{2}} \Gamma(m^* + 2 - \frac{nkp}{2})} \left(1 + \frac{\xi}{\gamma}\right)^{-(m^*+2)} \right] \\
&= \frac{\left(\frac{2m-kp}{2}\right) \left(\frac{2m-kp}{2} + 1\right)}{\left(\frac{2m+nkp-kp}{2} + 1\right)\gamma^2} \left[\frac{\Gamma(m^* + 2) \left(\frac{1}{\gamma}\right)^{\frac{nkp}{2}}}{\pi^{\frac{nkp}{2}} \Gamma(m^* + 2 - \frac{nkp}{2})} \left(1 + \frac{\xi}{\gamma}\right)^{-(m^*+2)} \right] \\
&= \frac{(2m - kp)(2m - kp + 2)}{2\gamma^2(2m + nkp - kp + 2)} g^{**}(\xi)
\end{aligned}$$

Siendo g^{**} la función que caracteriza a una distribución de Pearson Tipo VII con dimensión nkp y parámetros $\gamma^{**} = \gamma$, y $m^{**} = m^* + 2 = \frac{2m+nkp-kp}{2} + 2$. Es directo ver que:

$$E(y_i(W)) = \int_X u_i(W) |\Sigma|^{-\frac{n}{2}} \frac{(2m - kp)(2m - kp + 2)}{2\gamma^2(2m + nkp - kp + 2)} g^{**}(\xi) dX$$

lo cual, dado que las constantes salen de la integral, prueba la proposición.

QED

Lema IV.10. Sean B, C, D y X matrices de dimensión $p \times p$, entonces:

$$tr(BX'CXD) = (vec(X))'(B'D' \otimes C)vec(X) \quad (IV.27)$$

Demostración: se encuentra en [Neuenschwander y Flury, 2000, pag. 175].

Teorema IV.3. La matriz de información esperada de Fisher para $\underline{\beta}^*$ correspondiente a una distribución elíptica donde $D = [I_k \otimes \beta]'W[I_k \otimes \beta]$, es:

$$I_n(\underline{\beta}^*) = E \left(\frac{-\partial^2 \mathcal{L}(\Lambda, \beta)}{\partial \underline{\beta}^{*'} \partial \underline{\beta}^*} \right) \quad (IV.28)$$

donde

$$\frac{\partial^2 \mathcal{L}}{\partial \underline{\beta}^{*'} \partial \underline{\beta}^*} = \varrho^{-1} \left(\varrho^{-1} \left[\frac{g^{*2}(T)}{g^*(T)} - \frac{g^{*2}(T)}{g^{*2}(T)} \right] [AA'] + \left[\frac{g^{*'}(T)}{g^*(T)} \right] [B'] \right)$$

con

$$T = \varrho^{-1}(\text{tr}(\Lambda^{-1}D))$$

y donde

$$A = \left[\frac{\partial}{\partial \underline{\beta}^*} (\text{vec}(\beta))' \right] \sum_{i=1}^k \sum_{j=1}^k \left[(\Lambda^{ij} \otimes W_{ij}) + (\Lambda^{ij} \otimes W_{ji}) \right] \text{vec}(\beta)$$

y

$$B = \left[\frac{\partial}{\partial \underline{\beta}^*} \text{vec} \left(\left[\frac{\partial}{\partial \underline{\beta}^*} (\text{vec}(\beta))' \right] \right)' \right] \left(\left[\sum_{i=1}^k \sum_{j=1}^k \left[(\Lambda^{ij} \otimes W_{ij}) + (\Lambda^{ij} \otimes W_{ji}) \right] \text{vec}(\beta) \right] \otimes I_{\frac{p(p-1)}{2}} \right) \\ + \left[\frac{\partial}{\partial \underline{\beta}^*} (\text{vec}(\beta))' \right] \left(\sum_{i=1}^k \sum_{j=1}^k \left[(\Lambda^{ij} \otimes W_{ij}) + (\Lambda^{ij} \otimes W_{ji}) \right] \right)' \left[\frac{\partial}{\partial \underline{\beta}^*} (\text{vec}(\beta))' \right]'$$

Demostración:

La función log-verosimilitud es:

$$\mathcal{L} = C + \frac{n}{2} \log(|\Lambda^{-1}|) + \log[g^*(\varrho^{-1} \text{tr}(\Lambda^{-1}D))]$$

y se puede reescribir como:

$$\mathcal{L} = C + \frac{n}{2} \log(|\Lambda^{-1}|) + \log \left[g^* \left(\varrho^{-1} \left(\sum_{i=1}^k \sum_{j=1}^k \text{tr}(\Lambda^{ij} \beta' W_{ij} \beta) \right) \right) \right]$$

Por regla de la cadena se tiene que:

$$\frac{\partial \mathcal{L}}{\partial \underline{\beta}^*} = e^{-1} \left[\frac{g^{*'}(T)}{g^*(T)} \right] \left[\sum_{i=1}^k \sum_{j=1}^k \frac{\partial}{\partial \underline{\beta}^*} \text{tr}(\Lambda^{ij} \beta' W_{ij} \beta) \right]$$

con

$$T = e^{-1}(\text{tr}(\Lambda^{-1}D))$$

Por el **Lema IV.10** se puede reescribir a la traza como:

$$\text{tr}(\Lambda^{ij} \beta' W_{ij} \beta) = (\text{vec}(\beta))' (\Lambda^{ij} \otimes W_{ij}) \text{vec}(\beta)$$

De donde la log-verosimilitud es:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \underline{\beta}^*} &= e^{-1} \left[\frac{g^{*'}(T)}{g^*(T)} \right] \left[\sum_{i=1}^k \sum_{j=1}^k \frac{\partial}{\partial \underline{\beta}^*} \left((\text{vec}(\beta))' (\Lambda^{ij} \otimes W_{ij}) \text{vec}(\beta) \right) \right] \\ &= e^{-1} \left[\frac{g^{*'}(T)}{g^*(T)} \right] \left[\frac{\partial}{\partial \underline{\beta}^*} (\text{vec}(\beta))' \right] \sum_{i=1}^k \sum_{j=1}^k \left[(\Lambda^{ij} \otimes W_{ij}) + (\Lambda^{ij} \otimes W_{ji}) \right] \text{vec}(\beta) \\ &= e^{-1} \left[\frac{g^{*'}(T)}{g^*(T)} \right] [A] \end{aligned}$$

Aplicando nuevamente regla de la cadena se obtiene:

$$\frac{\partial^2 \mathcal{L}}{\partial \underline{\beta}^{*'} \partial \underline{\beta}^*} = (e^{-2}) \left[\frac{g^{*''}(T)}{g^*(T)} - \frac{g^{*'}(T)^2}{g^{*2}(T)} \right] [AA'] + (e^{-1}) \left[\frac{g^{*'}(T)}{g^*(T)} \right] \left[\frac{\partial}{\partial \underline{\beta}^{*'}} A \right]$$

Si se plantea que:

$$\begin{aligned} \frac{\partial}{\partial \underline{\beta}^{*'}} A &= \left[\frac{\partial}{\partial \underline{\beta}^*} A' \right]' \\ &= B' \end{aligned}$$

entonces aplicando la regla del producto se obtiene:

$$B = \left[\frac{\partial}{\partial \underline{\beta}^*} \text{vec} \left(\left[\frac{\partial}{\partial \underline{\beta}^*} (\text{vec}(\beta))' \right]' \right) \right] \left(\left[\sum_{i=1}^k \sum_{j=1}^k [(\Lambda^{ij} \otimes W_{ij}) + (\Lambda^{ji} \otimes W_{ji})] \text{vec}(\beta) \right] \otimes I_{\frac{p(p-1)}{2}} \right) \\ + \left[\frac{\partial}{\partial \underline{\beta}^*} (\text{vec}(\beta))' \right] \left(\sum_{i=1}^k \sum_{j=1}^k [(\Lambda^{ij} \otimes W_{ij}) + (\Lambda^{ji} \otimes W_{ji})] \right)' \left[\frac{\partial}{\partial \underline{\beta}^*} (\text{vec}(\beta))' \right]'$$

QED

Una vez mas se hace patente la imposibilidad de obtener una expresión cerrada para la matriz de información esperada de Fisher para $\underline{\beta}^*$ de manera general dentro de la familia de las distribuciones elípticas, por lo cual nuevamente se plantea la necesidad de conocer la función $g^*(t)$ para intentar obtener dichas expresiones de manera cerrada como en el caso Normal, donde $\left[\frac{g^{*''}(t)}{g^*(t)} - \frac{g^{*'}(t)^2}{g^{*2}(t)} \right] = 0$ y $\frac{g^{*'}(t)}{g^*(t)} = -\frac{1}{2}$ (ver expresión (II.31)). A continuación se plantea una vez más el caso de la distribución Pearson Tipo VII a manera de ejemplo.

Proposición IV.4. Sean x_1, \dots, x_n vectores aleatorios independientes con idéntica distribución Pearson Tipo VII $x_i \sim PP_p(\underline{\mu}, \gamma, m, \Sigma, g)$, la esperanza que aparece en el **Teorema IV.3** se puede obtener a partir de las esperanzas $E^{**}(WW')$ y $E^{***}(W)$, siendo $E^{**}(WW')$ la esperanza de WW' calculada para una distribución Pearson Tipo VII de dimensión nkp con parámetros $\gamma^{**} = \gamma$ y $m^{**} = \frac{2m+nkp-kp}{2} + 2$, y siendo $E^{***}(W)$ la esperanza de W calculada para una distribución Pearson Tipo VII de dimensión nkp con parámetros $\gamma^{***} = \gamma$ y $m^{***} = \frac{2m+nkp-kp}{2} + 1$. Ambas esperanzas puede calcularse utilizando el resultado del **Teorema III.5**.

Demostración:

Las esperanzas que aparecen en el **Teorema IV.3** tienen que ver con los términos $W_{ij}W_{kl}$ en el primer sumando, y con W_{ij} en el caso del segundo sumando. De hecho en el caso del primer sumando, dado que incluye la expresión $\left[\frac{g^{*''}(\xi)}{g^*(\xi)} - \frac{g^{*'}(\xi)^2}{g^{*2}(\xi)} \right]$, la prueba de que su esperanza se obtiene a partir de la esperanza $E^{**}(WW')$ es completamente análoga a la prueba de la **Proposición IV.3**. En el caso del segundo término se puede apreciar que aparece el término $\frac{g^{*'}(\xi)}{g^*(\xi)}$. De nuevo remitiéndose a la prueba de la **Proposición IV.3** se observa que al calcular la esperanza se tiene el término:

$$\begin{aligned}
g^*(\xi) \left[\frac{g^{**}(\xi)}{g^*(\xi)} \right] &= g^{**}(\xi) \\
&= \frac{\Gamma(m^*) \left(\frac{1}{\gamma}\right)^{\frac{nk\rho}{2}}}{\pi^{\frac{nk\rho}{2}} \Gamma(m^* - \frac{nk\rho}{2})} \left(\frac{-m^*}{\gamma}\right) \left(1 + \frac{\xi}{\gamma}\right)^{-m^*-1} \\
&= -\frac{m^* - \frac{nk\rho}{2}}{\gamma} \frac{\Gamma(m^* + 1) \left(\frac{1}{\gamma}\right)^{\frac{nk\rho}{2}}}{\pi^{\frac{nk\rho}{2}} \Gamma(m^* + 1 - \frac{nk\rho}{2})} \left(1 + \frac{\xi}{\gamma}\right)^{-(m^*+1)} \\
&= -\left(\frac{2m + k\rho}{2\gamma}\right) \frac{\Gamma(m^* + 1) \left(\frac{1}{\gamma}\right)^{\frac{nk\rho}{2}}}{\pi^{\frac{nk\rho}{2}} \Gamma(m^* + 1 - \frac{nk\rho}{2})} \left(1 + \frac{\xi}{\gamma}\right)^{-(m^*+1)} \\
&= -\left(\frac{2m + k\rho}{2\gamma}\right) g^{***}(\xi)
\end{aligned}$$

De donde se infiere que al obtener la esperanza del segundo sumando para el **Teorema IV.3**, las constantes salen y por tanto se tendrán esperanzas relacionadas con la distribución Pearson Tipo VII con dimensión $nk\rho$ y parámetros $\gamma^{***} = \gamma$ y $m^{***} = m^* + 1 = \frac{2m + nk\rho - k\rho}{2} + 1$.

QED

La obtención analítica de $E^{**}(WW')$ y $E^{***}(W)$ enunciadas en la **Proposición IV.4** no es sencilla. Una alternativa es encontrarlas numéricamente. Una segunda alternativa para obtener la estimación de la varianza del vector $\underline{\beta}^*$ consiste en calcular un estimador de la inversa de la matriz de información esperada de Fisher por medio de lo que se conoce como matriz de información observada de Fisher, que consiste en omitir el operador esperanza y sustituir los parámetros por sus correspondientes estimadores en la expresión (IV.28). Esta alternativa se justifica a partir de una comparación entre el uso de la matriz de información observada y esperada de Fisher. Efron y Hinkley muestran una justificación frecuentista para inclinar su preferencia por el uso de la primera (Efron & Hinkley, 1978). Para plantear esta propuesta se requiere de los dos resultados que a continuación se describen.

Lema IV.11. Sean B , X y C y matrices de dimensiones $m \times n$, $n \times l$ y $l \times r$ respectivamente, entonces:

$$vec(BXC) = (C' \otimes B)vec(X) \quad (IV.29)$$

Demostración: se encuentra en [Neuenschwander y Flury, 2000, pag. 175].

Lema IV.12. Sea $\beta = [\underline{\beta}_1, \dots, \underline{\beta}_p]$ una matriz ortogonal de dimensión $p \times p$, sean u_1, \dots, u_p escalares, y sea $U = bdiag[u_h I_p]_{h=1, \dots, p} = diag(u_1, \dots, u_p) \otimes I_p$. Si β_s^* es elemento de β^* un vector con $\frac{p(p-1)}{2}$ entradas funcionalmente independientes de β , entonces:

$$\left[\frac{\partial^2}{\partial \underline{\beta}^* \partial \beta_s^*} (vec(\beta))' \right] [U] [vec(\beta)] = - \left[\frac{\partial}{\partial \underline{\beta}^*} (vec(\beta))' \right] [U] \left[\frac{\partial}{\partial \beta_s^*} (vec(\beta)) \right] \quad (IV.30)$$

Demostración: se encuentra en [Neuenschwander y Flury, 2000, pag. 182].

Proposición IV.5. Un estimador de la matriz de información observada de Fisher para el vector $\underline{\beta}^*$ correspondiente al modelo del Análisis de Componentes Principales Comunes en Vectores Dependientes bajo el supuesto de distribución elíptica está dado por:

$$n \left[\frac{\partial}{\partial \underline{\hat{\beta}}^*} (vec(\hat{\beta}))' \right] (I_p \otimes \hat{\beta}) \left(\left[\sum_{i=1}^k \sum_{j=1}^k (\hat{\Lambda}^{ij} \otimes \hat{\Lambda}_{ij}) \right] - k I_{p^2} \right) (I_p \otimes \hat{\beta}') \left[\frac{\partial}{\partial \underline{\hat{\beta}}^*} (vec(\hat{\beta}))' \right]' \quad (IV.31)$$

Demostración:

Partiendo de los resultados del **Teorema IV.3**, se analiza primero el caso de \hat{B}' , para lo cual se recuerda que:

$$B = \left[\frac{\partial}{\partial \underline{\beta}^*} vec \left(\left[\frac{\partial}{\partial \underline{\beta}^*} (vec(\beta))' \right]' \right) \right] \left(\left[\sum_{i=1}^k \sum_{j=1}^k [(\Lambda^{ij} \otimes W_{ij}) + (\Lambda^{ij} \otimes W_{ji})] vec(\beta) \right] \otimes I_{\frac{p(p-1)}{2}} \right) \\ + \left[\frac{\partial}{\partial \underline{\beta}^*} (vec(\beta))' \right] \left(\sum_{i=1}^k \sum_{j=1}^k [(\Lambda^{ij} \otimes W_{ij}) + (\Lambda^{ij} \otimes W_{ji})] \right)' \left[\frac{\partial}{\partial \underline{\beta}^*} (vec(\beta))' \right]'$$

La dimensión de B es $\frac{p(p-1)}{2} \times \frac{p(p-1)}{2}$. Se tratará a continuación de analizar por separado las columnas del primer sumando de $B = B_1 + B_2$, esto es:

$$B_{1s} = \left[\frac{\partial^2}{\partial \underline{\beta}^* \partial \beta_s^*} (vec(\beta))' \right] \left(\sum_{i=1}^k \sum_{j=1}^k [(\Lambda^{ij} \otimes W_{ij}) + (\Lambda^{ij} \otimes W_{ji})] vec(\beta) \right)$$

Y por lo tanto:

$$\hat{B}_{1s} = \left[\frac{\partial^2}{\partial \underline{\hat{\beta}}^* \partial \hat{\beta}_s^*} (vec(\hat{\beta}))' \right] \left(\sum_{i=1}^k \sum_{j=1}^k [(\hat{\Lambda}^{ij} \otimes W_{ij}) + (\hat{\Lambda}^{ij} \otimes W_{ji})] vec(\hat{\beta}) \right)$$

Recordando que $W_{ij} = u^{-1} \underline{e}^{-1} \widehat{\Sigma}_{ij} = u^{-1} \widehat{\beta} \widehat{\Lambda}_{ij} \widehat{\beta}'$, se tiene:

$$\begin{aligned} \widehat{B}_{1s} &= \left[\frac{\partial^2}{\partial \widehat{\beta}_* \partial \widehat{\beta}_s} (\text{vec}(\widehat{\beta}))' \right] \left(\sum_{i=1}^k \sum_{j=1}^k [(\widehat{\Lambda}^{ij} \otimes u^{-1} \widehat{\beta} \widehat{\Lambda}_{ij} \widehat{\beta}') + (\widehat{\Lambda}^{ij} \otimes u^{-1} \widehat{\beta} \widehat{\Lambda}_{ij} \widehat{\beta}')] \text{vec}(\widehat{\beta}) \right) \\ &= \left[\frac{\partial^2}{\partial \widehat{\beta}_* \partial \widehat{\beta}_s} (\text{vec}(\widehat{\beta}))' \right] \left(2 \sum_{i=1}^k \sum_{j=1}^k (\widehat{\Lambda}^{ij} \otimes u^{-1} \widehat{\beta} \widehat{\Lambda}_{ij} \widehat{\beta}') \text{vec}(\widehat{\beta}) \right) \end{aligned}$$

Utilizando el **Lema IV.11** se tiene:

$$\begin{aligned} \widehat{B}_{1s} &= \left[\frac{\partial^2}{\partial \widehat{\beta}_* \partial \widehat{\beta}_s} (\text{vec}(\widehat{\beta}))' \right] \left(2 \sum_{i=1}^k \sum_{j=1}^k \text{vec} (u^{-1} \widehat{\beta} \widehat{\Lambda}_{ij} \widehat{\beta}' \widehat{\Lambda}^{ij}) \right) \\ &= \left[\frac{\partial^2}{\partial \widehat{\beta}_* \partial \widehat{\beta}_s} (\text{vec}(\widehat{\beta}))' \right] \left(2 \sum_{i=1}^k \sum_{j=1}^k \text{vec} (u^{-1} \widehat{\beta} \widehat{\Lambda}_{ij} \widehat{\Lambda}^{ij}) \right) \end{aligned}$$

Utilizando una vez más el **Lema IV.11** se tiene:

$$\widehat{B}_{1s} = 2u^{-1} \left[\frac{\partial^2}{\partial \widehat{\beta}_* \partial \widehat{\beta}_s} (\text{vec}(\widehat{\beta}))' \right] \left(\sum_{i=1}^k \sum_{j=1}^k (\widehat{\Lambda}_{ij} \widehat{\Lambda}^{ij} \otimes I_p) \text{vec}(\widehat{\beta}) \right)$$

Como $\sum_{i=1}^k \sum_{j=1}^k \widehat{\Lambda}_{ij} \widehat{\Lambda}^{ij} = kl_p$ (pues $\widehat{\Lambda}_{ii} \widehat{\Lambda}^{ii} = I_p$ y $\widehat{\Lambda}_{ij} \widehat{\Lambda}^{ij} = 0$ para $i \neq j$), entonces se tiene que:

$$\sum_{i=1}^k \sum_{j=1}^k \widehat{\Lambda}_{ij} \widehat{\Lambda}^{ij} \otimes I_p = kl_p \otimes I_p$$

por lo que:

$$\widehat{B}_{1s} = 2u^{-1} \left[\frac{\partial^2}{\partial \widehat{\beta}_* \partial \widehat{\beta}_s} (\text{vec}(\widehat{\beta}))' \right] [kl_p \otimes I_p] \text{vec}(\widehat{\beta})$$

Pero $kl_p \otimes I_p$ tiene la forma de U del **Lema IV.12**, por lo tanto:

$$\widehat{B}_{1s} = -2u^{-1} \left[\frac{\partial}{\partial \widehat{\beta}_*} (\text{vec}(\widehat{\beta}))' \right] [kl_p \otimes I_p] \left[\frac{\partial}{\partial \widehat{\beta}_s} (\text{vec}(\widehat{\beta})) \right]$$

Conjuntando las columnas, y utilizando el hecho de que $kl_p \otimes I_p = kl_{p^2}$ se llega a:

$$\begin{aligned}\hat{B}_1 &= -2u^{-1} \left[\frac{\partial}{\partial \underline{\hat{\beta}}^*} (\text{vec}(\hat{\beta}))' \right] [kI_{p^2}] \left[\frac{\partial}{\partial \underline{\hat{\beta}}^{*'}} (\text{vec}(\hat{\beta})) \right] \\ &= -2u^{-1} \left[\frac{\partial}{\partial \underline{\hat{\beta}}^*} (\text{vec}(\hat{\beta}))' \right] [kI_{p^2}] \left[\frac{\partial}{\partial \underline{\hat{\beta}}^*} (\text{vec}(\hat{\beta}))' \right]'\end{aligned}$$

Recordando que el segundo sumando \hat{B}_2 es:

$$\begin{aligned}\hat{B}_2 &= \left[\frac{\partial}{\partial \underline{\hat{\beta}}^*} (\text{vec}(\hat{\beta}))' \right] \left(\sum_{i=1}^k \sum_{j=1}^k [(\hat{\Lambda}^{ij} \otimes W_{ij}) + (\hat{\Lambda}^{ij} \otimes W_{ji})] \right)' \left[\frac{\partial}{\partial \underline{\hat{\beta}}^*} (\text{vec}(\hat{\beta}))' \right]' \\ &= 2u^{-1} \left[\frac{\partial}{\partial \underline{\hat{\beta}}^*} (\text{vec}(\hat{\beta}))' \right] \left(\sum_{i=1}^k \sum_{j=1}^k [\hat{\Lambda}^{ij} \otimes \hat{\beta} \hat{\Lambda}_{ij} \hat{\beta}'] \right) \left[\frac{\partial}{\partial \underline{\hat{\beta}}^*} (\text{vec}(\hat{\beta}))' \right]'\end{aligned}$$

Por tanto $\hat{B} = \hat{B}_1 + \hat{B}_2$ es:

$$\hat{B} = 2u^{-1} \left[\frac{\partial}{\partial \underline{\hat{\beta}}^*} (\text{vec}(\hat{\beta}))' \right] \left(\left[\sum_{i=1}^k \sum_{j=1}^k (\hat{\Lambda}^{ij} \otimes \hat{\beta} \hat{\Lambda}_{ij} \hat{\beta}') \right] - kI_{p^2} \right) \left[\frac{\partial}{\partial \underline{\hat{\beta}}^*} (\text{vec}(\hat{\beta}))' \right]'$$

Se puede demostrar, al realizar las operaciones, que:

$$\hat{\Lambda}^{ij} \otimes \hat{\beta} \hat{\Lambda}_{ij} \hat{\beta}' = (I_p \otimes \hat{\beta}) (\hat{\Lambda}^{ij} \otimes \hat{\Lambda}_{ij}) (I_p \otimes \hat{\beta}')$$

y que:

$$(I_p \otimes \hat{\beta}) kI_{p^2} (I_p \otimes \hat{\beta}') = kI_{p^2}$$

Con lo cual:

$$\hat{B}' = 2u^{-1} \left[\frac{\partial}{\partial \underline{\hat{\beta}}^*} (\text{vec}(\hat{\beta}))' \right] (I_p \otimes \hat{\beta}) \left(\left[\sum_{i=1}^k \sum_{j=1}^k (\hat{\Lambda}^{ij} \otimes \hat{\Lambda}_{ij}) \right] - kI_{p^2} \right) (I_p \otimes \hat{\beta}') \left[\frac{\partial}{\partial \underline{\hat{\beta}}^*} (\text{vec}(\hat{\beta}))' \right]'$$

Por otra parte, recordando que A tiene la siguiente forma:

$$A = \left[\frac{\partial}{\partial \underline{\beta}^*} (\text{vec}(\beta))' \right] \sum_{i=1}^k \sum_{j=1}^k [(\Lambda^{ij} \otimes W_{ij}) + (\Lambda^{ij} \otimes W_{ji})] \text{vec}(\beta)$$

Se puede apreciar entonces que \hat{A} tiene una estructura similar a \hat{B}_{1s} por lo que se tiene que:

$$\begin{aligned}\hat{A} &= 2u^{-1} \left[\frac{\partial}{\partial \underline{\hat{\beta}}^*} (\text{vec}(\hat{\beta}))' \right] [kI_{p^2}] \text{vec}(\hat{\beta}) \\ &= 2u^{-1}k \left[\frac{\partial}{\partial \underline{\hat{\beta}}^*} (\text{vec}(\hat{\beta}))' \right] \text{vec}(\hat{\beta})\end{aligned}$$

Debido a las restricciones de ortogonalidad impuestas sobre $\hat{\beta}$, se tiene que:

$$\left[\frac{\partial}{\partial \underline{\hat{\beta}}^*} (\text{vec}(\hat{\beta}))' \right] \text{vec}(\hat{\beta}) = 0$$

Ahora bien, dado que $\hat{D} = [I_k \otimes \hat{\beta}]'W[I_k \otimes \hat{\beta}]$ se tiene:

$$\begin{aligned}\text{tr}(\hat{\Lambda}^{-1}\hat{D}) &= \text{tr}(\hat{\Lambda}^{-1}[I_k \otimes \hat{\beta}]'W[I_k \otimes \hat{\beta}]) \\ &= \text{tr}(\hat{\Lambda}^{-1}[I_k \otimes \hat{\beta}]'u^{-1}[I_k \otimes \hat{\beta}]\hat{\Lambda}[I_k \otimes \hat{\beta}][I_k \otimes \hat{\beta}]) \\ &= u^{-1}\text{tr}(\hat{\Lambda}^{-1}\hat{\Lambda}) \\ &= u^{-1}\text{tr}(I_{kp}) \\ &= u^{-1}kp\end{aligned}$$

De donde se concluye que:

$$\begin{aligned}\hat{T} &= \varrho^{-1}\text{tr}(\hat{\Lambda}^{-1}\hat{D}) \\ &= \varrho^{-1}u^{-1}kp\end{aligned}$$

$$I_n(\hat{\beta}^*) = -2u^{-1}\varrho^{-1} \left[\frac{g^{*'}(\varrho^{-1}u^{-1}kp)}{g^*(\varrho^{-1}u^{-1}kp)} \right] C$$

donde

$$C = \left[\frac{\partial}{\partial \underline{\hat{\beta}}^*} (\text{vec}(\hat{\beta}))' \right] (I_p \otimes \hat{\beta}) \left(\left[\sum_{i=1}^k \sum_{j=1}^k (\hat{\Lambda}^{ij} \otimes \hat{\Lambda}_{ij}) \right] - kI_{p^2} \right) (I_p \otimes \hat{\beta}') \left[\frac{\partial}{\partial \underline{\hat{\beta}}^*} (\text{vec}(\hat{\beta}))' \right]'$$

Dado que $\varrho^{-1} = \frac{E(R^{*2})}{nkp}$, se tiene entonces:

$$\begin{aligned} I_n(\widehat{\underline{\beta}}^*) &= -2u^{-1} \frac{E(R^{*2})}{nkp} \left[\frac{g^{*'} \left(\frac{E(R^{*2})}{nkp} \left(\frac{kp}{u} \right) \right)}{g^* \left(\frac{E(R^{*2})}{nkp} \left(\frac{kp}{u} \right) \right)} \right] C \\ &= nu^{-1} \frac{E(R^{*2})}{nkp} \left[\frac{-\frac{2}{n} g^{*'} \left(\frac{E(R^{*2})}{nu} \right)}{g^* \left(\frac{E(R^{*2})}{nu} \right)} \right] C \end{aligned}$$

Se identifica al término $u = \frac{E(R^{*2})}{nkp} \left[\frac{-\frac{2}{n} g^{*'} \left(\frac{E(R^{*2})}{nu} \right)}{g^* \left(\frac{E(R^{*2})}{nu} \right)} \right]$, de donde:

$$\begin{aligned} I_n(\widehat{\underline{\beta}}^*) &= nu^{-1} uC \\ &= nC \end{aligned}$$

QED

IV.3 Comparaciones Teóricas con Respecto a los Resultados Bajo el Supuesto de Normalidad

Teorema IV.4. En el modelo definido para el Análisis de Componentes Principales Comunes en Vectores Dependientes bajo el supuesto de distribución elíptica, obtener los estimadores de Λ y β por máxima verosimilitud utilizando el **Teorema IV.1**, es equivalente a obtenerlos encontrando primero el estimador por máxima verosimilitud $\widehat{\Psi}$ para Ψ , y luego diagonalizando simultáneamente los bloques que lo integran en la forma $\widehat{\Psi} = [\widehat{\Psi}]_{ij}, i, j = 1, \dots, k$, utilizando para ello el **Teorema II.7**.

Demostración:

La diferencia entre el sistema determinado por el **Teorema II.7** y el correspondiente al **Teorema IV.1** es la presencia, en éste segundo, del término u .

Para determinar la equivalencia basta con probar que para Ψ , su estimador por máxima verosimilitud $\widehat{\Psi}$ es tal que $\widehat{\Psi}_{ij} = uW_{ij}$, ya que en ese caso los sistemas de ambos teoremas serán iguales.

El **Corolario III.1** establece que el estimador por máxima verosimilitud para Ψ en este caso es:

$$\hat{\Psi} = \frac{E(R^{*2})}{nkp} \delta_{\max(g)} W$$

por lo cual para que $\hat{\Psi}_{ij} = uW_{ij}$ es necesario probar que

$$u = \frac{E(R^{*2})}{nkp} \delta_{\max(g)}$$

pero

$$u = \frac{E(R^{*2})}{nkp} \left[\frac{-\frac{2}{n} g^{*'} \left(\frac{E(R^{*2})}{nu} \right)}{g^* \left(\frac{E(R^{*2})}{nu} \right)} \right]$$

Si se define $y = \frac{nkp}{E(R^{*2})} u$, de donde $u = \frac{E(R^{*2})}{nkp} y$ y se sustituye a u en la ecuación anterior se obtiene:

$$y = \left(\frac{-2}{n} \right) \frac{g^{*'} \left(\frac{kp}{y} \right)}{g^* \left(\frac{kp}{y} \right)}$$

Esta expresión para encontrar y es exactamente la misma expresión definida por el **Lema III.1** para encontrar $\delta_{\max(g)}$ en el caso en que se desea estimar por máxima verosimilitud a Ψ , por tanto

$$y = \delta_{\max(g)}$$

y esto prueba que

$$u = \frac{E(R^{*2})}{nkp} \delta_{\max(g)}$$

QED

Teorema IV.5. El estimador máximo verosímil para el parámetro β en el modelo de Componentes Principales Comunes en Vectores Dependientes es el mismo para cualquier distribución elíptica de la que provengan los datos.

Demostración:

El **Corolario III.1** establece que para las distribuciones elípticas el estimador máximo verosímil de Ψ es $\hat{\Psi} = \frac{E(R^{*2})}{np} \delta_{max(g)} W$ donde $\frac{E(R^{*2})}{np} \delta_{max(g)}$ es un valor que dependerá de los parámetros de la distribución que se utilice y del tamaño de muestra, que para efectos prácticos son constantes. Se puede plantear entonces que $\hat{\Psi} = uW$ donde u es una constante que en el caso de la distribución normal tiene valor de $\frac{1}{n}$.

El **Teorema IV.4** establece que para encontrar los estimadores máximo verosímiles de β y Λ basta aplicar el **Teorema II.7** a $\hat{\Psi} = uW$. Los estimadores resultantes son tales que:

$$\hat{\Psi} = [I_k \otimes \hat{\beta}] \hat{\Lambda} [I_k \otimes \hat{\beta}]'$$

es decir:

$$uW = [I_k \otimes \hat{\beta}] \hat{\Lambda} [I_k \otimes \hat{\beta}]'$$

expresión que, en el caso de normalidad, puede ser reescrita como:

$$\frac{1}{n}W = [I_k \otimes \hat{\beta}] \hat{\Lambda}^* [I_k \otimes \hat{\beta}]'$$

de donde se tiene que: $\hat{\Lambda} = nu\hat{\Lambda}^*$.

Es claro que al aplicar el **Teorema II.3** se obtendrán los valores de $\hat{\beta}$ y $\hat{\Lambda}^*$, y que para encontrar la matriz $\hat{\Lambda}$ correspondiente a cierta distribución elíptica bastaría multiplicar $\hat{\Lambda}^*$ por nu correspondiente a la distribución elíptica supuesta. Sin embargo no sería necesario modificar el estimador de $\hat{\beta}$. Esto prueba que el estimador de máxima verosimilitud $\hat{\beta}$ tiene la misma expresión para cualquier distribución elíptica de la que provengan los datos (en este sentido es robusto).

QED

Teorema IV.6. La distribución asintótica estimada mediante la información observada de Fisher, para el parámetro β del modelo de Componentes Principales Comunes en Vectores Dependientes es la misma para cualquier distribución elíptica de la que provengan los datos.

Demostración: se deriva de manera directa de la **Proposición IV.5**.

IV.4 Comparaciones Empíricas con Respecto a los Resultados Bajo el Supuesto de Normalidad

En esta sección se presentan tres técnicas de estimación de varianza que pueden utilizarse como alternativas al método de estimación asintótica. En esta sección se muestran las versiones de las técnicas, aplicables al caso en que se dispone de una muestra aleatoria simple, con el objetivo principal de mostrar su efectividad, para lo cual se simulan dos ejemplos con grandes tamaños de muestra, en los cuales se realizan comparaciones con la correspondiente estimación asintótica. Se presenta también su aplicación en dos ejemplos reales con tamaños de muestra pequeños. Las técnicas que se presentarán tienen la característica de poder ser adaptados al caso en que se dispone de muestras obtenidas a partir de diseños muestrales complejos, donde el método de estimación asintótica puede ser impreciso al no considerar el efecto del diseño muestral. En ello radica la importancia de validar su efectividad.

Las técnicas que a continuación se describen se conocen como métodos de remuestreo, y se basan en la idea de generar k submuestras (de tamaño similar) a partir de la muestra original, para cada una de las cuales se obtiene el estimador de un parámetro de interés, y con dichos estimadores se calcula un estimador de varianza.

Estas técnicas parten del concepto de distribución muestral de un estimador, que establece que cuando se conoce al total de la población y se pueden generar todas las submuestras posibles de un tamaño establecido, es posible entonces encontrar la distribución del estimador de un parámetro de interés (y particularmente su varianza). En este sentido, los métodos de remuestreo generan una aproximación a la varianza de los estimadores, la cual será más precisa en tanto el tamaño de la muestra disponible sea mayor. Esto implica que si bien estos métodos pueden ser alternativos al asintótico, en muestras pequeñas se puede tener estimaciones poco precisas.

En la práctica, los métodos de remuestreo se utilizan generalmente cuando la muestra ya se ha obtenido, es decir, no se puede modificar el tamaño de muestra. Por lo cual, la precisión en el estimador depende principalmente del número k de submuestras que se generan. En particular, dado que se trata de replicar la idea de la distribución de muestreo, las submuestras generadas deben ser tales que constituyan una muestra aleatoria del total de submuestras del mismo tamaño que se hubieran podido obtener para toda la población. Este punto se discutirá posteriormente.

IV.4.1 El Método de Grupos Aleatorios

Definición IV.3. El método de Grupos Aleatorios para estimación de varianza consiste en asignar aleatoriamente a cada individuo de la muestra original, en una y sólo una de las k submuestras, con lo cual dichas submuestras no tienen elementos en común. Para cada submuestra se obtiene un estimador $\hat{\theta}_i$ del parámetro de interés, calculando el estimador de varianza de la siguiente forma:

$$\widehat{V}(\hat{\theta}) = \frac{1}{k(k-1)} \sum_{i=1}^k (\hat{\theta}_i - \hat{\theta})^2$$

donde:

$$\hat{\theta} = \sum_{i=1}^k \frac{\hat{\theta}_i}{k}$$

Es posible demostrar que $\widehat{V}(\hat{\theta})$ es un estimador sesgado, y que su sesgo va a tender a ser positivo y a reducirse en muestras de gran tamaño. (Wolter, 1985).

IV.4.2 El Método de Jackknife

Definición IV.4. El método de Jackknife para estimación de varianza consiste en dividir aleatoriamente la muestra original en k subconjuntos sin elementos en común, y obtener a partir de dichos subconjuntos, todas las posibles uniones distintas de $k-1$ grupos cada vez. Dichas uniones constituyen las k submuestras a partir de cada una de las cuales se obtiene un estimador $\hat{\theta}_i$ del parámetro de interés, calculando el estimador de varianza de la siguiente forma:

$$\widehat{V}(\hat{\theta}) = \frac{k-1}{k} \sum_{i=1}^k (\hat{\theta}_i - \hat{\theta})^2$$

donde $\hat{\theta}$ es el estimador para la muestra original.

El estimador de la varianza obtenido por el método de Jackknife es sesgado, pero su sesgo es de menor orden que el correspondiente al método de grupos aleatorios, y también tiende a reducirse en muestras de gran tamaño. (Wolter, 1985).

IV.4.3 El Método de Replicaciones (Balanceadas)

Definición IV.5. El método de Replicaciones para estimación de varianza consiste en dividir aleatoriamente la muestra original en k subconjuntos sin elementos en común. Se selecciona aleatoriamente un porcentaje fijo (frecuentemente se recurre a la mitad) de los k subconjuntos, y esa selección se une para formar la primera submuestra, replicándose el procedimiento hasta obtener k submuestras. Para cada submuestra se obtiene un estimador $\hat{\theta}_i$ del parámetro de interés, calculando el estimador de varianza de la siguiente forma:

$$V(\hat{\theta}) = \frac{1}{k} \sum_{i=1}^k (\hat{\theta}_i - \hat{\theta})^2$$

donde $\hat{\theta}$ es el estimador para la muestra original.

Para reducir el sesgo del estimador, se ha propuesto “balancear” las replicaciones, de manera que todos los s grupos estén representados en la misma proporción en las submuestras. (Wolter, 1985).

IV.4.4 Estabilidad de los Métodos de Estimación de Varianza

La precisión de los métodos presentados depende del tamaño de la muestra original, y del número de subgrupos (k) que se elija utilizar. La elección de un número óptimo de subgrupos depende de dos factores fundamentalmente, el primero es la precisión que se desea obtener, la cual puede ser medida a través del coeficiente de variación del estimador de varianza, buscando que éste sea mínimo, o por medio de métodos estadísticos alternativos. El segundo es el costo computacional (traducido a costo en tiempo) que se tiene. Por ello, la elección óptima del número de subgrupos debe realizarse mediante una medida conjunta de la precisión y el costo computacional (Wolter, 1985).

En el contexto del presente proyecto se puede elegir el número óptimo de grupos, evaluando la diferencia entre los estimadores de varianza obtenidos por los métodos presentados, con respecto a la varianza asintótica de los estimadores. Esto se justifica partiendo del supuesto de que, a mayor tamaño de muestra, el estimador asintótico será más preciso, con lo cual es válido tomarlo como punto de comparación en ese caso.

Ahora bien, dado que los estimadores que se tratan en el presente proyecto, son vectores o matrices, la aplicación de los métodos se realizará obteniendo la varianza estimada de los componentes de los vectores (o matrices), es decir, obteniendo la varianza estimada de las entradas de los mismos para posteriormente generar vectores o matrices de varianzas estimadas. Esto implica, por una parte, que para evaluar la diferencia entre los estimadores de varianza se requiere evaluar diferencias matriciales, para lo cual se requiere de medidas de disimilaridad de matrices. Por otra parte implica que sólo pueden obtenerse las matrices de varianzas de las entradas de los vectores o matrices de estimadores, pero no las correspondientes covarianzas entre dichas entradas.

Existen medidas de disimilaridad de matrices generalmente basadas en la idea de medir la distancia de la matriz con respecto a la singularidad. Sin embargo en el contexto de este proyecto, estas medidas no son de utilidad. En el área de estudio de imágenes digitales, se han desarrollado medidas de disimilaridad de matrices con una mayor cercanía al contexto referido, en la búsqueda de comparación de imágenes observadas con imágenes reales (por ejemplo para imágenes de satélite). Dichas metodologías aplican el concepto de matriz a los píxeles de las referidas imágenes. Este campo está en desarrollo, como se puede apreciar en artículos recientes al respecto (Ortega, et. al., 2006). Para este proyecto se utilizará una medida simple de disimilaridad, con la idea de plantear la posibilidad, posteriormente, de estudiar alternativas más complejas (y más precisas). La medida a utilizar será el promedio de las distancias (diferencias al cuadrado) entre entradas, de las matrices de varianzas obtenidas por los métodos previamente presentados, con respecto a la matriz de varianzas asintótica de los estimadores.

IV.4.5 Aplicación de los Métodos de Estimación de Varianza a datos simulados

A continuación se presentan cuatro ejemplos que muestran el desempeño de los métodos de estimación de varianza descritos en la sección anterior. Los primeros dos ejemplos son datos simulados con tamaños de muestra grande, por lo cual se comparan los métodos presentados con la varianza asintótica de los elementos de la matriz β estimada para el modelo de Componentes Principales Comunes en Vectores Dependientes bajo el supuesto de normalidad, haciendo uso de la **Proposición IV.5**. También se comparan con los que se obtienen al suponer distribución Pearson Tipo VII. Para ello se generaron programas en Fortran que permiten simular, para una matriz de varianzas y covarianzas hipotética, valores con distribución Normal y con distribución Pearson Tipo VII partiendo de los resultados presentados en el **Capítulo III**. Se incluyen otros dos ejemplos reales, con tamaños de muestras pequeños, para evaluar el comportamiento de los métodos de estimación de varianza en ese caso.

Ejemplo IV.1. Este ejemplo no incluye datos reales, está tomado de un artículo que lo utiliza para mostrar la eficiencia del Algoritmo FG+ al plantear el uso de matrices mal condicionadas (cercanas a la singularidad) (Neuenschwander y Flury, 1994). Se plantea un caso hipotético en el que existieran dos grupos dependientes, con dos variables en cada grupo, siendo la matriz de varianzas y covarianzas:

$$V(\underline{x}) = \begin{bmatrix} 2.246 & -2.257 & 2.135 & -1.204 \\ -2.257 & 3.869 & -3.219 & 0.618 \\ 2.135 & -3.219 & 3.175 & -1.264 \\ -1.204 & 0.618 & -1.264 & 3.080 \end{bmatrix}$$

Se parte primero del supuesto de Normalidad, y utilizando el Algoritmo FG+ ortogonal se encuentra que la estimación de la Matriz β es:

$$\hat{\beta} = \begin{bmatrix} 0.7300 & 0.6835 \\ -0.6835 & 0.7300 \end{bmatrix}$$

La estimación asintótica de las varianzas de los elementos de la Matriz β , obtenida utilizando la matriz de información de Fisher, es:

$$V(\hat{\beta})_{Asintot} = \begin{bmatrix} 2.10E-06 & 2.40E-06 \\ 2.40E-06 & 2.10E-06 \end{bmatrix}$$

Para evaluar el desempeño de los métodos de estimación de varianza, se simulan 10,000 observaciones bajo el supuesto de normalidad. Se obtienen los estimadores de las varianzas de los elementos de la Matriz β para los métodos de Replicaciones Repetidas, Grupos Aleatorios y Jackknife. Se analizan los resultados para distinto número de subconjuntos k que van de 10 a 100 en múltiplos de 10, de 100 a 500 en múltiplos de 50, buscando evaluar la consistencia en las estimaciones, y por medio de la medida de disimilaridad descrita en la sección anterior, seleccionando el número de subconjuntos que genera la estimación más cercana a la correspondiente estimación asintótica de las varianzas de la Matriz β , esto bajo la premisa de que al tener un gran número de observaciones, la estimación asintótica es precisa.

Se identificó que para el método de Replicaciones Repetidas, el número de subconjuntos que genera la estimación más cercana a la correspondiente estimación asintótica es $k = 60$, lo cual implica haber generado 60 submuestras de tamaño aproximado 167. Se puede apreciar la similitud que tiene dicha estimación con la correspondiente estimación asintótica. Para el método de Grupos Aleatorios se identificó que el número de subconjuntos que genera la estimación más cercana a la correspondiente estimación asintótica es $k = 30$, lo cual implica haber generado 30 submuestras de tamaño aproximado 5,000. Se puede apreciar nuevamente

la similitud que tiene dicha estimación con la correspondiente estimación asintótica. Finalmente, para el método de Jackknife se identificó que el número de subconjuntos que genera la estimación más cercana a la correspondiente estimación asintótica es $k = 30$, lo cual implica haber generado 30 submuestras de tamaño aproximado 9,667. También se observa una similitud con la correspondiente estimación asintótica.

$$\widehat{V(\hat{\beta})}_{Replica(60)} = \begin{bmatrix} 1.77E-06 & 2.00E-06 \\ 2.00E-06 & 1.77E-06 \end{bmatrix}$$

$$\widehat{V(\hat{\beta})}_{G.Aleat(30)} = \begin{bmatrix} 1.97E-06 & 2.20E-06 \\ 2.20E-06 & 1.97E-06 \end{bmatrix}$$

$$\widehat{V(\hat{\beta})}_{Jackkn(30)} = \begin{bmatrix} 1.85E-06 & 2.09E-06 \\ 2.09E-06 & 1.85E-06 \end{bmatrix}$$

Para evaluar el cambio en el desempeño de los métodos, y mostrar empíricamente la validez de los resultados teóricos previos, se repite el análisis anterior pero planteando ahora el supuesto de distribución Pearson Tipo VII con parámetros $\gamma = 15$ y $m = 5$ (elegidos al azar) y simulando 10,000 observaciones. Se puede apreciar que la estimación de la Matriz β se mantiene prácticamente sin cambios, y lo mismo ocurre con la estimación asintótica de las varianzas de los elementos de la Matriz $\hat{\beta}$, validando empíricamente los **Teoremas IV.5 y IV.6**.

$$\hat{\beta} = \begin{bmatrix} 0.7299 & 0.6836 \\ -0.6836 & 0.7299 \end{bmatrix}$$

$$\widehat{V(\hat{\beta})}_{Asintot} = \begin{bmatrix} 2.10E-06 & 2.40E-06 \\ 2.40E-06 & 2.10E-06 \end{bmatrix}$$

En el caso del supuesto de distribución Pearson Tipo VII ocurre que, aún cuando se selecciona el número de subconjuntos que genera la estimación más cercana a la correspondiente estimación asintótica, dichas estimaciones difieren entre sí.

$$\widehat{V(\hat{\beta})}_{Replica(70)} = \begin{bmatrix} 3.19E-06 & 3.58E-06 \\ 3.58E-06 & 3.19E-06 \end{bmatrix}$$

$$\widehat{V(\hat{\beta})}_{G.Aleat(40)} = \begin{bmatrix} 3.78E-06 & 4.32E-06 \\ 4.32E-06 & 3.78E-06 \end{bmatrix}$$

$$\widehat{V(\hat{\beta})}_{Jackkn(40)} = \begin{bmatrix} 3.66E-06 & 4.12E-06 \\ 4.12E-06 & 3.66E-06 \end{bmatrix}$$

Para evaluar de manera general la estabilidad de los métodos, se generó para el caso de la distribución Normal Multivariada y para los número de subconjuntos que genera la estimación más cercana a la correspondiente estimación asintótica para cada caso, la estimación de la matriz de varianzas para hasta 100,000 vectores simulados, presentándose algunos cortes para evaluar el comportamiento de los estimadores.

Para presentar de manera sintética los resultados obtenidos, se generó para cada uno de ellos la medida de disimilaridad con respecto a la correspondiente estimación asintótica. La medida utilizada fué la raíz cuadrada de la suma de las diferencias al cuadrado de cada una de las entradas de las matrices, multiplicada por el número de vectores simulados hasta el corte respectivo. Esta última multiplicación se realiza para poder comparar la medida calculada para distinto número de simulaciones, dado que, como se puede observar a partir de la estimación asintótica, las varianzas estimadas dependen inversamente del número de datos para los cuales se estima.

***Medida de disimilaridad de la matriz de varianzas estimada
por método con respecto a la estimación asintótica***

Datos simulados	Replicaciones	Grupos Aleatorios	Jackknife
100	74.8527	2.4890	195.1790
500	0.0087	4.0760	0.0012
1000	0.0141	8.2305	0.0012
2000	0.0171	14.4850	0.0081
4000	0.0005	9.5670	0.0096
6000	0.0079	14.2710	0.0004
8000	0.0186	0.0008	0.0047
10000	0.0043	0.0034	0.0057
20000	0.0007	0.0168	0.0139
50000	0.0092	0.0107	0.0105
100000	0.0015	0.0064	0.0067

Se puede observar que con pocas observaciones, el número de subconjuntos elegido no genera adecuadas estimaciones, lo cual implica una mayor necesidad de elegir adecuadamente el número de grupos en muestras pequeñas (más adelante se presentan ejemplos con ese caso), siendo este elemento particularmente importante en el método de Grupos Aleatorios, para el que se observa muy poca precisión incluso con muestras de alrededor de 6,000 datos. Conforme el tamaño de muestra aumenta, los tres estimadores se vuelven más precisos.

Ejemplo IV.2. Este ejemplo tampoco incluye datos reales, y fué planteado por sus autores (Neuenschwander y Flury, 1994) buscando evaluar el comportamiento del Algoritmo FG+ cuando de antemano se sabe que existe un vector propio común, en un caso hipotético en el que existieran dos grupos dependientes, con cuatro variables en cada grupo, siendo la matriz de varianzas y covarianzas:

$$V(\underline{x}) = \begin{bmatrix} 90.40 & -32.80 & -27.20 & 9.60 & 0.00 & -30.00 & -10.00 & 0.00 \\ -32.80 & 109.60 & -9.60 & -27.20 & -30.00 & 0.00 & 0.00 & -10.00 \\ -27.20 & -9.60 & 109.60 & -32.80 & -10.00 & 0.00 & 0.00 & -30.00 \\ 9.60 & -27.20 & -32.80 & 90.40 & 0.00 & -10.00 & -30.00 & 0.00 \\ 0.00 & -30.00 & -10.00 & 0.00 & 109.60 & 30.40 & 12.80 & 7.20 \\ -30.00 & 0.00 & 0.00 & -10.00 & 30.40 & 109.60 & 7.20 & 12.80 \\ -10.00 & 0.00 & 0.00 & -30.00 & 12.80 & 7.20 & 90.40 & 49.60 \\ 0.00 & -10.00 & -30.00 & 0.00 & 7.20 & 12.80 & 49.60 & 90.40 \end{bmatrix}$$

La estimación de la Matriz β es:

$$\hat{\beta} = \begin{bmatrix} 0.5000 & -0.1922 & 0.6498 & -0.5393 \\ 0.5000 & 0.1811 & -0.7118 & -0.4588 \\ 0.5001 & 0.6874 & 0.2168 & 0.4799 \\ 0.4999 & -0.6765 & -0.1549 & 0.5181 \end{bmatrix}$$

La estimación asintótica de las varianzas de los elementos de la Matriz β es:

$$\widehat{V(\hat{\beta})}_{Asintot} = \begin{bmatrix} 3.19E-05 & 6.68E-05 & 5.45E-05 & 8.16E-05 \\ 2.84E-05 & 7.48E-05 & 4.20E-05 & 9.40E-05 \\ 2.35E-05 & 1.71E-05 & 1.01E-04 & 4.06E-05 \\ 2.54E-05 & 1.54E-05 & 1.05E-04 & 3.67E-05 \end{bmatrix}$$

Se replica el esquema de análisis del ejemplo anterior buscando los número de subconjuntos que genera la estimación más cercana a la correspondiente estimación asintótica, primero bajo el supuesto de normalidad simulando 10, 000 observaciones.

$$\widehat{V(\hat{\beta})}_{Replica(30)} = \begin{bmatrix} 2.40E-05 & 8.41E-05 & 8.26E-05 & 1.18E-04 \\ 3.63E-05 & 1.20E-04 & 1.07E-04 & 2.11E-04 \\ 2.64E-05 & 2.22E-05 & 7.85E-05 & 7.78E-05 \\ 3.71E-05 & 1.63E-05 & 2.53E-04 & 5.58E-05 \end{bmatrix}$$

$$\widehat{V(\hat{\beta})}_{G.Alcat(500)} = \begin{bmatrix} 4.17E-05 & 1.03E-04 & 1.12E-03 & 7.35E-04 \\ 1.78E-04 & 6.69E-05 & 2.14E-03 & 1.24E-03 \\ 1.68E-04 & 3.39E-04 & 1.82E-04 & 3.45E-04 \\ 6.72E-05 & 3.73E-04 & 1.65E-04 & 7.80E-04 \end{bmatrix}$$

$$\widehat{V(\hat{\beta})}_{Jackkn(90)} = \begin{bmatrix} 2.82E-05 & 1.14E-04 & 7.40E-05 & 8.45E-05 \\ 2.85E-05 & 1.21E-04 & 7.95E-05 & 1.51E-04 \\ 1.92E-05 & 1.82E-05 & 1.32E-04 & 4.12E-05 \\ 3.15E-05 & 2.26E-05 & 1.89E-04 & 4.90E-05 \end{bmatrix}$$

Las medidas de disimilaridad con respecto a la estimación asintótica son **0.00046**, **0.0821** y **0.00017** respectivamente. De los resultados anteriores se puede apreciar que en el Método de Grupos Aleatorios genera la estimación menos precisa en este caso.

Cuando se analizan los datos simulando **10,000** observaciones bajo el supuesto de distribución Pearson Tipo VII con parámetros $\gamma = 15$ y $m = 10$, se aprecia que tanto el Método de Replicaciones Repetidas genera estimaciones muy poco precisas. En este caso las medidas de disimilaridad con respecto a la estimación asintótica son **78.16**, **0.0876** y **0.00073** respectivamente. Para el método de Jackknife no existen diferencias tan notorias con respecto a la estimación asintótica como en el ejemplo anterior.

$$\widehat{V(\hat{\beta})}_{Replica(30)} = \begin{bmatrix} 1.23E-04 & 9.26E-05 & 5.58E-02 & 2.03E-04 \\ 1.74E-05 & 8.98E-05 & 6.83E-02 & 2.32E-04 \\ 2.82E-05 & 2.79E-05 & 6.26E-03 & 9.59E-05 \\ 7.07E-05 & 3.82E-05 & 3.56E-03 & 6.62E-05 \end{bmatrix}$$

$$\widehat{V(\hat{\beta})}_{G.Aleat(500)} = \begin{bmatrix} 4.74E-05 & 7.81E-05 & 1.02E-03 & 8.85E-04 \\ 1.85E-04 & 6.19E-05 & 2.28E-03 & 1.26E-03 \\ 1.47E-04 & 2.33E-04 & 2.30E-04 & 2.93E-04 \\ 7.58E-05 & 2.78E-04 & 1.63E-04 & 7.97E-04 \end{bmatrix}$$

$$\widehat{V(\hat{\beta})}_{Jackkn(40)} = \begin{bmatrix} 5.24E-05 & 1.22E-04 & 1.32E-04 & 1.83E-04 \\ 2.02E-05 & 9.97E-05 & 1.14E-04 & 2.42E-04 \\ 2.05E-05 & 1.67E-05 & 1.55E-04 & 7.76E-05 \\ 3.50E-05 & 2.41E-05 & 2.51E-04 & 3.33E-05 \end{bmatrix}$$

Ejemplo IV.3. Este ejemplo contiene datos reales de medidas de audiometría tomados a **100** hombres de la misma edad, midiendo pérdida de audición en decibeles, para frecuencias de 500Hz, 1000Hz y 2000Hz al oído izquierdo y al oído derecho (dos grupos) (Jackson, 1991; Flury y Neuenschwander, 1995), siendo la matriz de varianzas y covarianzas:

$$V(\underline{x}) = \begin{bmatrix} 41.07 & 37.73 & 28.13 & 31.79 & 26.30 & 14.12 \\ 37.73 & 57.32 & 44.44 & 29.75 & 34.24 & 25.30 \\ 28.13 & 44.44 & 119.70 & 18.64 & 31.21 & 71.26 \\ 31.79 & 29.75 & 18.64 & 50.75 & 30.23 & 10.52 \\ 26.30 & 34.24 & 31.21 & 30.23 & 40.92 & 24.62 \\ 14.12 & 25.30 & 71.26 & 10.52 & 24.62 & 86.30 \end{bmatrix}$$

La estimación de la Matriz β es:

$$\hat{\beta} = \begin{bmatrix} 0.6767 & 0.6573 & 0.3316 \\ -0.7248 & 0.5157 & 0.4569 \\ 0.1293 & -0.5495 & 0.8254 \end{bmatrix}$$

La estimación asintótica de las varianzas de los elementos de la Matriz β es:

$$\widehat{V(\hat{\beta})}_{Asintot} = \begin{bmatrix} 1.04E-03 & 1.67E-03 & 2.80E-03 \\ 7.30E-04 & 2.40E-03 & 1.88E-03 \\ 1.11E-03 & 4.00E-03 & 1.77E-03 \end{bmatrix}$$

Debido a que el número de observaciones es 100, este ejemplo servirá para mostrar el desempeño de los métodos no asintóticos en muestras pequeñas. Dado que no se dispone de los datos, se simulan las 100 observaciones de una población con matriz de varianzas y covarianzas igual a la muestral, primero bajo el supuesto de normalidad. Si bien esta simulación conlleva un margen de error por simular los datos, permite observar si general todos los métodos generan estimaciones similares. El número de subconjuntos en cada caso se elige arbitrariamente, tratando de hacer comparables los métodos. En este caso seleccionar en el método de grupos aleatorios $k = 20$ implica generar 20 submuestras de tamaño 50. En Grupos aleatorios la elección de $k = 2$ implica generar 2 submuestras de tamaño 50. Finalmente en Jackknife elegir $k = 50$ implica generar 50 submuestras de tamaño 98.

$$\widehat{V(\hat{\beta})}_{Replica(20)} = \begin{bmatrix} 9.01E-04 & 4.32E-03 & 4.42E-02 \\ 7.00E-04 & 2.33E-03 & 3.70E-02 \\ 3.46E-04 & 7.98E-02 & 3.83E-03 \end{bmatrix}$$

$$\widehat{V(\hat{\beta})}_{G.Aleat(2)} = \begin{bmatrix} 7.26E-04 & 3.03E-03 & 4.20E-03 \\ 4.27E-04 & 4.27E-04 & 2.72E-03 \\ 3.11E-04 & 7.72E-03 & 2.81E-03 \end{bmatrix}$$

$$\widehat{V(\hat{\beta})}_{Jackkn(50)} = \begin{bmatrix} 9.80E-04 & 1.54E-03 & 3.17E-03 \\ 6.65E-04 & 2.09E-03 & 1.39E-03 \\ 7.58E-04 & 4.03E-03 & 1.58E-03 \end{bmatrix}$$

Si se supone distribución Pearson Tipo VII con parámetros $\gamma = 15$ y $m = 10$, y se simulan 100 observaciones, no se observan resultados muy similares al caso normal.

$$\widehat{V(\hat{\beta})}_{Replica(20)} = \begin{bmatrix} 1.57E-03 & 2.09E-03 & 9.79E-04 \\ 1.19E-03 & 2.77E-03 & 1.68E-03 \\ 1.53E-03 & 2.28E-03 & 1.04E-03 \end{bmatrix}$$

$$\widehat{V(\hat{\beta})}_{G.Aleat(2)} = \begin{bmatrix} 2.17E-04 & 9.78E-05 & 1.57E-03 \\ 1.15E-04 & 2.09E-03 & 1.63E-03 \\ 3.47E-04 & 2.89E-03 & 1.28E-03 \end{bmatrix}$$

$$\widehat{V(\hat{\beta})}_{Jackkn(50)} = \begin{bmatrix} 1.58E-03 & 2.55E-03 & 2.02E-03 \\ 1.09E-03 & 2.34E-03 & 1.93E-03 \\ 1.36E-03 & 3.66E-03 & 1.50E-03 \end{bmatrix}$$

De los resultados anteriores se concluye que en muestras pequeñas no es claro que los métodos no asintóticos sean útiles ya que los resultados tienen diferencias entre métodos. En este caso tampoco se puede suponer que la estimación asintótica sea válida. Es entonces importante estudiar más detenidamente estos métodos, principalmente en lo que se refiere a la elección del número de subconjuntos óptimo, lo cual es una línea de investigación de interés que puede ser continuada posteriormente. A continuación se analiza otro ejemplo con muestras pequeñas, pero donde sí se dispone de los datos.

Ejemplo IV.4. Este ejemplo contiene también datos reales de medidas realizadas al femur y la tibia de las patas traseras derechas de 88 insectos denominados water strider, efectuadas en tres etapas distintas de su crecimiento (Flury y Neuenschwander, 1995; Klingenberg, Neuenschwander y Flury, 1996).

En este caso se dispone de los datos, y la estimación de la Matriz β es:

$$\hat{\beta} = \begin{bmatrix} 0.7761 & 0.6307 \\ -0.6307 & 0.7761 \end{bmatrix}$$

La estimación asintótica de las varianzas de los elementos de la Matriz β es:

$$\widehat{V(\hat{\beta})}_{Asintot} = \begin{bmatrix} 3.03E-04 & 4.59E-04 \\ 4.59E-04 & 3.03E-04 \end{bmatrix}$$

El **Teorema IV.5** implica que es indistinto suponer distribución Normal que distribución Pearson Tipo VII para encontrar los estimadores no asintóticos de las varianzas de los elementos de la Matriz β , que en este caso son:

$$\widehat{V(\hat{\beta})}_{Replica(20)} = \begin{bmatrix} 4.44E-04 & 7.35E-04 \\ 7.35E-04 & 4.44E-04 \end{bmatrix}$$

$$\widehat{V(\hat{\beta})}_{G.Aleat(2)} = \begin{bmatrix} 6.10E-04 & 9.40E-04 \\ 9.40E-04 & 6.10E-04 \end{bmatrix}$$

$$\widehat{V(\hat{\beta})}_{Jackkn(50)} = \begin{bmatrix} 5.96E-04 & 9.25E-04 \\ 9.25E-04 & 5.96E-04 \end{bmatrix}$$

Si bien estos resultados son más similares entre sí que en el ejemplo anterior, es difícil concluir que sean precisos en la estimación de varianza buscada. Nuevamente se reitera la necesidad de estudiar más a detalle los casos de muestras pequeñas, y en general el problema de la determinación del número óptimo de subconjuntos a elegir en cada método.

Capítulo V. Análisis de Componentes Principales Comunes en Vectores Dependientes Bajo el Supuesto de Distribución Elíptica en Diseños Muestrales Complejos

En este capítulo se presentará una forma de estimar parámetros de interés cuando se supone Distribución Elíptica sobre un conjunto de datos que han sido obtenidos mediante muestras de diseño complejo. En particular se muestra la forma de obtener los parámetros del modelo de Componentes Principales Comunes en Vectores Dependientes, así como la manera de realizar inferencias sobre los mismos.

Se utiliza la terminología planteada por los textos de muestreo, donde se destaca la utilidad práctica que tiene el uso de estratificaciones o conglomeraciones para generar muestras de menor tamaño que generen los mismos niveles de confiabilidad para las estimaciones que se obtengan a partir de ellas. Particularmente se plantea el uso de muestreos probabilísticos en varias etapas de selección, siendo los elementos de la primera etapa de selección denotados como unidades primarias de muestreo (UPM), los elementos de la segunda etapa son denominados unidades secundarias de muestreo (USM) y así sucesivamente.

Estos muestreos complejos se caracterizan por asignar distintas probabilidades de selección a los individuos, debido justamente a las conglomeraciones o estratificaciones que se realizan, por lo cual cada individuo tiene un peso distinto. No tomar en cuenta el hecho de que se ha procedido de manera distinta a un muestreo aleatorio simple implica generar estimaciones sobre una población distinta a la que se desea estudiar, ya que la muestra obtenida no contiene observaciones independientes ni idénticamente distribuidas, aún cuando provengan de la misma población.

V.1 Estimación Puntual de los Parámetros por el Método de Pseudo Máxima Verosimilitud

Una de las estrategias para considerar el hecho de que las observaciones ya no son idénticamente distribuidas, consiste en suponer distribuciones nominales para cada observación, que reflejen el hecho de que su selección no fué realizada por el método aleatorio simple o irrestricto. Una alternativa para considerar este tipo de distribuciones nominales es la técnica denominada Pseudo Máxima Verosimilitud, que a continuación se define.

Definición V.1. Sean $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$ un conjunto de vectores que provienen de una población con distribución $f(\underline{x}; \theta)$, seleccionados por un esquema de diseño muestral complejo, el método de Pseudo Máxima Verosimilitud consiste en proponer $f_i(\underline{x}_i; \theta)$ una función de distribución nominal para cada vector \underline{x}_i , y estimar θ con el valor de $\tilde{\theta}$ que maximiza:

$$n^{-1} \sum_{i=1}^n \log f_i(\underline{x}_i; \tilde{\theta})$$

En la literatura referida al tema se menciona que la “mejor” propuesta para la distribución nominal $f_i(\underline{x}_i; \theta)$ debe corresponder a la “mejor” aproximación de entre todas las posibles propuestas (Skinner, et. al., 1989). En el caso de tratar de encontrar un estimador que contemple el diseño muestral complejo se tiene que proponer esa “mejor” aproximación desde alguna perspectiva sustentada. A continuación se presenta una serie de resultados que permitirán establecer una propuesta sustentada para las distribuciones nominales $f_i(\underline{x}_i; \theta)$ correspondientes al caso en que los datos provienen de una población con Distribución Normal Multivariada, con la posibilidad de generalizarlos al caso de Distribuciones Elípticas.

Definición V.2. Sean $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$ un conjunto de vectores que provienen de una población con distribución $f(\underline{x}; \theta)$, seleccionados por un esquema de diseño muestral complejo. Sean $\omega_1, \omega_2, \dots, \omega_n$ los factores de expansión correspondientes de modo que $\omega_i = P_i^{-1}$, donde P_i es la probabilidad de selección de \underline{x}_i . Nótese que $\sum_{i=1}^n \omega_i = N$ (Donde es N el tamaño de la población). Sean $\alpha_1, \alpha_2, \dots, \alpha_n$ definidos como $\alpha_i = \omega_i \left(\frac{n}{N}\right)$ de modo que $\sum_{i=1}^n \alpha_i = n$. Se define a la Log-Verosimilitud Ponderada como:

$$\mathcal{L}^* = \log[L^*] = \sum_{i=1}^n \alpha_i \log f(\underline{x}_i; \theta)$$

Esta definición parte de la idea de que la información que cada unidad de muestreo proporciona, no es igual, sino que depende justamente del peso que tiene en la muestra, y por ello se plantea la opción de ponderar dicha información de manera proporcional a los ponderadores originales. Además, es claro que bajo muestreo aleatorio simple, esta Log-Verosimilitud Ponderada coincide con la Log-Verosimilitud tradicional ya que en ese caso $\alpha_i = 1 \forall i$.

A continuación se muestra una serie de desarrollos que permitirán plantear la obtención de estimadores a partir de la maximización de la función de Log-Verosimilitud Ponderada para el caso particular de la distribución Normal Multivariada. Se utiliza una perspectiva matricial.

Proposición V.1. Sean $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$ vectores que provienen de una población con distribución Normal Multivariada $\underline{x}_i \sim N_p(\underline{\mu}, \underline{\Sigma})$, seleccionados por un esquema de diseño muestral complejo, entonces la Log-Verosimilitud Ponderada es igual a:

$$\mathcal{L}^* = \log[L^*] = -\frac{n}{2} \log |\underline{\Sigma}| - \frac{np}{2} \log(2\pi) - \frac{1}{2} \left[(\text{vec}(X') - \underline{1} \otimes \underline{\mu})' (d \otimes \underline{\Sigma})^{-1} (\text{vec}(X') - \underline{1} \otimes \underline{\mu}) \right]$$

donde $d = \text{diag} \left(\frac{1}{\alpha_1}, \frac{1}{\alpha_2}, \dots, \frac{1}{\alpha_n} \right)$, con $\underline{1}$ y X definidos en la **Sección III.2**.

Demostración:

$$\begin{aligned} \log[L^*] &= \sum_{i=1}^n \alpha_i \log f(\underline{x}_i) \\ &= \sum_{i=1}^n \alpha_i \log \left[|\underline{\Sigma}|^{-\frac{1}{2}} \left(\frac{1}{(2\pi)^{\frac{p}{2}}} \right) \exp \left(-\frac{1}{2} (\underline{x}_i - \underline{\mu})' \underline{\Sigma}^{-1} (\underline{x}_i - \underline{\mu}) \right) \right] \\ &= \sum_{i=1}^n \alpha_i \left[\log |\underline{\Sigma}|^{-\frac{1}{2}} + \log \left(\frac{1}{(2\pi)^{\frac{p}{2}}} \right) + \log \left(\exp \left(-\frac{1}{2} (\underline{x}_i - \underline{\mu})' \underline{\Sigma}^{-1} (\underline{x}_i - \underline{\mu}) \right) \right) \right] \\ &= \sum_{i=1}^n \alpha_i \left[-\frac{1}{2} \log |\underline{\Sigma}| - \frac{p}{2} \log(2\pi) - \frac{1}{2} (\underline{x}_i - \underline{\mu})' \underline{\Sigma}^{-1} (\underline{x}_i - \underline{\mu}) \right] \\ &= \sum_{i=1}^n \alpha_i \left[-\frac{1}{2} \log |\underline{\Sigma}| - \frac{p}{2} \log(2\pi) \right] - \frac{1}{2} \sum_{i=1}^n \alpha_i (\underline{x}_i - \underline{\mu})' \underline{\Sigma}^{-1} (\underline{x}_i - \underline{\mu}) \\ &= n \left[-\frac{1}{2} \log |\underline{\Sigma}| - \frac{p}{2} \log(2\pi) \right] - \frac{1}{2} \sum_{i=1}^n (\underline{x}_i - \underline{\mu})' \left(\frac{\underline{\Sigma}}{\alpha_i} \right)^{-1} (\underline{x}_i - \underline{\mu}) \\ &= -\frac{n}{2} \log |\underline{\Sigma}| - \frac{np}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^n (\underline{x}_i - \underline{\mu})' \left(\frac{\underline{\Sigma}}{\alpha_i} \right)^{-1} (\underline{x}_i - \underline{\mu}) \\ &= -\frac{n}{2} \log |\underline{\Sigma}| - \frac{np}{2} \log(2\pi) - \frac{1}{2} \left[(\text{vec}(X') - \underline{1} \otimes \underline{\mu})' (d \otimes \underline{\Sigma})^{-1} (\text{vec}(X') - \underline{1} \otimes \underline{\mu}) \right] \end{aligned}$$

QED

Proposición V.2. Sea $\underline{z} \sim N_{np}(\underline{1} \otimes \underline{\mu}, d \otimes \underline{\Sigma})$, entonces su Log-Verosimilitud es:

$$\log[L] = -\frac{p}{2} \log \left(\prod_{i=1}^n \left(\frac{1}{\alpha_i} \right) \right) - \frac{n}{2} \log |\underline{\Sigma}| - \frac{np}{2} \log(2\pi) - \frac{1}{2} \left[(\underline{z} - \underline{1} \otimes \underline{\mu})' (d \otimes \underline{\Sigma})^{-1} (\underline{z} - \underline{1} \otimes \underline{\mu}) \right]$$

Demostración:

$$\begin{aligned}
\log[L] &= \log f(z) \\
&= \log \left[|d \otimes \Sigma|^{-\frac{1}{2}} \left(\frac{1}{(2\pi)^{\frac{p}{2}}} \right) \exp \left(-\frac{1}{2} (z - \underline{1} \otimes \underline{\mu})' d \otimes \Sigma^{-1} (z - \underline{1} \otimes \underline{\mu}) \right) \right] \\
&= \log |d \otimes \Sigma|^{-\frac{1}{2}} + \log \left(\frac{1}{(2\pi)^{\frac{p}{2}}} \right) + \log \left(\exp \left(-\frac{1}{2} (z - \underline{1} \otimes \underline{\mu})' d \otimes \Sigma^{-1} (z - \underline{1} \otimes \underline{\mu}) \right) \right) \\
&= -\frac{1}{2} \log |d \otimes \Sigma| - \frac{p}{2} \log(2\pi) - \frac{1}{2} (z - \underline{1} \otimes \underline{\mu})' d \otimes \Sigma^{-1} (z - \underline{1} \otimes \underline{\mu}) \\
&= -\frac{1}{2} \log (|d|^p |\Sigma|^n) - \frac{p}{2} \log(2\pi) - \frac{1}{2} (z - \underline{1} \otimes \underline{\mu})' d \otimes \Sigma^{-1} (z - \underline{1} \otimes \underline{\mu}) \\
&= -\frac{1}{2} \log |d|^p - \frac{1}{2} \log |\Sigma|^n - \frac{p}{2} \log(2\pi) - \frac{1}{2} (z - \underline{1} \otimes \underline{\mu})' d \otimes \Sigma^{-1} (z - \underline{1} \otimes \underline{\mu}) \\
&= -\frac{p}{2} \log \left(\prod_{i=1}^n \left(\frac{1}{\alpha_i} \right) \right) - \frac{n}{2} \log |\Sigma| - \frac{p}{2} \log(2\pi) - \frac{1}{2} (z - \underline{1} \otimes \underline{\mu})' d \otimes \Sigma^{-1} (z - \underline{1} \otimes \underline{\mu})
\end{aligned}$$

QED

Proposición V.3. La Log-Verosimilitud Ponderada planteada en la **Proposición V.1** es equivalente a la Log-Verosimilitud de la **Proposición V.2**.

Demostración:

Es claro ver que para $z = \text{vec}(X')$, se tiene que $\log[L] = \log[L^*] + \log k$ siendo:

$$k = \left(\prod_{i=1}^n \left(\frac{1}{\alpha_i} \right) \right)^{-\frac{p}{2}}$$

Por lo tanto $\log[L] = \log[k * L^*]$ de donde $L = k * L^*$, que es justamente la definición de equivalencia entre verosimilitudes.

QED

El resultado de la **Proposición V.3** establece la equivalencia entre realizar inferencias a partir de la Log-Verosimilitud de $\text{vec}(X') \sim N_{np}(\underline{1} \otimes \underline{\mu}, d \otimes \Sigma)$ con respecto a realizarlas a partir de la Log-Verosimilitud Ponderada de x_1, x_2, \dots, x_n vectores que provienen de una población con distribución Normal Multivariada $x_j \sim N_p(\underline{\mu}, \Sigma)$, seleccionados por un esquema

de diseño muestral complejo. Esto es de utilidad puesto que ya se conoce la forma de realizar inferencias utilizando la función de Log-Verosimilitud de un vector, pudiendo utilizarse para ello las herramientas previamente descritas en el **Capítulo III** concernientes a trabajar con la distribución de una matriz, que como se mencionó en su momento, puede simplificar algunos desarrollos teóricos.

Proposición V.4. Sea $\underline{z} \sim N_{np}(\underline{1} \otimes \underline{\mu}, d \otimes \underline{\Sigma})$ con $d = \text{diag}\left(\frac{1}{\alpha_1}, \frac{1}{\alpha_2}, \dots, \frac{1}{\alpha_n}\right)$, entonces se tiene que la distribución marginal de cada vector \underline{z}_i de dimensión p tal que $\underline{z}' = (\underline{z}'_1 \underline{z}'_2 \dots \underline{z}'_n)$ es tal que:

$$\underline{z}_i \sim N_p\left(\underline{\mu}, \frac{\underline{\Sigma}}{\alpha_i}\right)$$

Demostración:

Dado que la distribución Normal Multivariada pertenece a la familia de las distribuciones elípticas, se puede utilizar el **Teorema III.1** para encontrar la distribución marginal de \underline{z}_1 . Para ello se requiere obtener primero la distribución esférica correspondiente. Si se define a A de modo que $AA' = d \otimes \underline{\Sigma}$, entonces el vector $\underline{y} = A^{-1}(\underline{z} - \underline{1} \otimes \underline{\mu})$ es tal que $\underline{y} \sim N_{np}(\mathbf{0}, I)$.

Se aplica el **Teorema III.1** al vector \underline{y} para encontrar la función $g_{\underline{y}_1}(t)$ que caracteriza al vector marginal \underline{y}_1 de la siguiente forma:

$$g_{\underline{y}_1}(t) = \frac{\pi^{\frac{np-p}{2}}}{\Gamma\left(\frac{np-p}{2}\right)} \int_t^\infty (y-t)^{\frac{np-p}{2}-1} \frac{1}{(2\pi)^{\frac{np}{2}}} \exp\left(-\frac{y}{2}\right) dy$$

Se utiliza la transformación $s = y - t$, de donde $ds = dy$ con $s \in (0, \infty)$, entonces:

$$\begin{aligned} g_{\underline{y}_1}(t) &= \frac{\pi^{\frac{np-p}{2}}}{\Gamma\left(\frac{np-p}{2}\right) (2\pi)^{\frac{np}{2}}} \int_0^\infty (s)^{\frac{np-p}{2}-1} \exp\left(-\frac{s+t}{2}\right) ds \\ &= \left[\frac{\pi^{\frac{np-p}{2}}}{(2\pi)^{\frac{np}{2}}} \right] \exp\left(-\frac{t}{2}\right) \left[2^{\frac{np-p}{2}} \right] \int_0^\infty \left[\frac{\left(\frac{1}{2}\right)^{\frac{np-p}{2}}}{\Gamma\left(\frac{np-p}{2}\right)} \right] s^{\frac{np-p}{2}-1} \exp\left(-\frac{s}{2}\right) ds \end{aligned}$$

De donde se identifica dentro de la integral a una distribución *Gamma* $\left(\frac{np-p}{2}, \frac{1}{2}\right)$, es decir, una distribución $\chi^2_{(np-p)}$, por lo cual el valor de dicha integral es uno, teniéndose entonces:

$$\begin{aligned}
g_{y_1}(t) &= \left[\frac{\pi^{\frac{np-p}{2}}}{(2\pi)^{\frac{np}{2}}} \right] \exp\left(-\frac{t}{2}\right) \left[2^{\frac{np-p}{2}} \right] \\
&= \left[\frac{1}{(2\pi)^{\frac{p}{2}}} \right] \exp\left(-\frac{t}{2}\right)
\end{aligned}$$

que corresponde a una distribución Normal p-variada, esto es, $y_1 \sim N_p(\mathbf{0}, I)$.

Para determinar la distribución de z_1 , se tiene que si B es tal que $B'B = \Sigma$, entonces se puede probar que $A = d^{\frac{1}{2}} \otimes B$ con $d^{\frac{1}{2}} = \text{diag}\left(\frac{1}{\sqrt{\alpha_1}}, \frac{1}{\sqrt{\alpha_2}}, \dots, \frac{1}{\sqrt{\alpha_n}}\right)$, ya que por propiedades del producto de kronecker:

$$\begin{aligned}
[d^{\frac{1}{2}} \otimes B]'[d^{\frac{1}{2}} \otimes B] &= (d^{\frac{1}{2}'} d^{\frac{1}{2}}) \otimes (B'B) \\
&= d \otimes \Sigma \\
&= A'A
\end{aligned}$$

Este resultado implica que A es una matriz diagonal por bloques, por lo que se puede obtener la distribución de z_1 utilizando el primer bloque de dicha matriz de la forma $z_1 = A_1 y_1 + \underline{\mu}$, con lo cual

$$z_1 \sim N_p\left(\underline{\mu}, \frac{\Sigma}{\alpha_1}\right)$$

Obtener la distribución marginal del resto de los vectores z_j es análogo, intercambiando de lugar dichos vectores, lo cual es posible ya que por la forma de la varianza (diagonal por bloques) se deduce que cada vector z_j es no correlacionado con el resto, y debido a que se está estudiando a la distribución Normal Multivariada, en la cual es válida la implicación de independencia a partir de no correlación, entonces los vectores referidos son permutables.

QED

Con esta serie de resultados se puede demostrar el siguiente teorema, estableciendo una propuesta sustentada para las distribuciones nominales requeridas en el método de Pseudo Máxima Verosimilitud para el caso en que los datos provienen de una población con distribución Normal Multivariada

Teorema V.1. Sean $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$ un conjunto de vectores que provienen de una población con distribución Normal Multivariada $\underline{x}_i \sim N_p(\underline{\mu}, \underline{\Sigma})$, seleccionados por un esquema de diseño muestral complejo. Sean $\omega_1, \omega_2, \dots, \omega_n$ los factores de expansión correspondientes de modo que $\omega_i = P_i^{-1}$, donde P_i es la probabilidad de selección de \underline{x}_i . Sean $\alpha_1, \alpha_2, \dots, \alpha_n$ definidos como $\alpha_i = \omega_i \left(\frac{n}{N}\right)$. Entonces las inferencias a partir de la Log-Verosimilitud Ponderada de la **Definición V.2** son equivalentes a las obtenidas a partir del método de Pseudo Máxima Verosimilitud de la **Definición V.1**, utilizando $\underline{x}_i \sim N_p\left(\underline{\mu}, \frac{\underline{\Sigma}}{\alpha_i}\right)$ como la función de distribución nominal para cada vector \underline{x}_i .

Demostración:

Ya se demostró previamente que la función de Log-Verosimilitud Ponderada para el caso de la distribución Normal Multivariada es:

$$\log[L^*] = -\frac{n}{2} \log |\underline{\Sigma}| - \frac{np}{2} \log (2\pi) - \frac{1}{2} \sum_{i=1}^n (\underline{x}_i - \underline{\mu})' \left(\frac{\underline{\Sigma}}{\alpha_i}\right)^{-1} (\underline{x}_i - \underline{\mu})$$

Por su parte, la función de Pseudo Verosimilitud utilizando las distribuciones nominales $\underline{x}_i \sim N_p\left(\underline{\mu}, \frac{\underline{\Sigma}}{\alpha_i}\right)$ es:

$$\begin{aligned} PV &= n^{-1} \sum_{i=1}^n \log \left[\left| \frac{\underline{\Sigma}}{\alpha_i} \right|^{-\frac{1}{2}} \left(\frac{1}{2\pi}\right)^{\frac{p}{2}} \exp \left(-\frac{1}{2} (\underline{x}_i - \underline{\mu})' \left(\frac{\underline{\Sigma}}{\alpha_i}\right)^{-1} (\underline{x}_i - \underline{\mu}) \right) \right] \\ &= n^{-1} \sum_{i=1}^n \left[-\frac{1}{2} \log \left| \frac{\underline{\Sigma}}{\alpha_i} \right| - \frac{p}{2} \log (2\pi) - \frac{1}{2} (\underline{x}_i - \underline{\mu})' \left(\frac{\underline{\Sigma}}{\alpha_i}\right)^{-1} (\underline{x}_i - \underline{\mu}) \right] \\ &= n^{-1} \sum_{i=1}^n \left[-\frac{1}{2} \log (\alpha_i^{-n} |\underline{\Sigma}|) - \frac{p}{2} \log (2\pi) - \frac{1}{2} (\underline{x}_i - \underline{\mu})' \left(\frac{\underline{\Sigma}}{\alpha_i}\right)^{-1} (\underline{x}_i - \underline{\mu}) \right] \\ &= n^{-1} \sum_{i=1}^n \left[\frac{n}{2} \log (\alpha_i) - \frac{1}{2} \log |\underline{\Sigma}| - \frac{p}{2} \log (2\pi) - \frac{1}{2} (\underline{x}_i - \underline{\mu})' \left(\frac{\underline{\Sigma}}{\alpha_i}\right)^{-1} (\underline{x}_i - \underline{\mu}) \right] \\ &= n^{-1} \left[\left(\sum_{i=1}^n \frac{n}{2} \log (\alpha_i) \right) - \frac{n}{2} \log |\underline{\Sigma}| - \frac{np}{2} \log (2\pi) - \frac{1}{2} \sum_{i=1}^n (\underline{x}_i - \underline{\mu})' \left(\frac{\underline{\Sigma}}{\alpha_i}\right)^{-1} (\underline{x}_i - \underline{\mu}) \right] \end{aligned}$$

Las funciones $\log[L^*]$ y nPV difieren por el sumando $\sum_{i=1}^n \frac{n}{2} \log(\alpha_i)$, que es constante, por lo que ambas alcanzan el valor óptimo en términos de $\underline{\mu}$ y $\underline{\Sigma}$ en los mismos puntos, es decir que las inferencias serán equivalentes.

QED

Este teorema establece el hecho de que, en el caso de la distribución Normal Multivariada, el método de estimación basado en maximizar la Log-Verosimilitud Ponderada es un método de Pseudo Máxima Verosimilitud, y define además la forma de las distribuciones nominales en dicho caso. A partir del teorema también se establecen las bases de la posible generalización para ser aplicado a distribuciones de la familia elíptica en general.

Proposición V.5. Sean $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$ vectores que provienen de una población con distribución Elíptica $\underline{x}_i \sim EC_p(\underline{\mu}, \underline{\Sigma}, g)$, seleccionados por un esquema de diseño muestral complejo. Sean α_i definidos como en el **Teorema V.1**. Entonces las distribuciones nominales definidas por $\underline{x}_i \sim EC_p(\underline{\mu}, \frac{\underline{\Sigma}}{\alpha_i}, g)$ pueden ser utilizadas para obtener estimadores por el método de Pseudo Máxima Verosimilitud para los parámetros $\underline{\mu}$ y $\underline{\Sigma}$.

Demostración: Es directa debido a que las distribuciones nominales planteadas satisfacen los requisitos de la **Definición V.1**.

Proposición V.6. Sean $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$ vectores que provienen de una población con distribución Elíptica $\underline{x}_i \sim EC_p(\underline{\mu}, \underline{\Sigma}, g)$, seleccionados por un esquema de diseño muestral complejo. Sea \underline{z} un vector aleatorio tal que $\underline{z} \sim EC_{np}(\underline{1} \otimes \underline{\mu}, d \otimes \underline{\Sigma}, g^*)$ con $d = \text{diag}\left(\frac{1}{\alpha_1}, \frac{1}{\alpha_2}, \dots, \frac{1}{\alpha_n}\right)$, para α_i definidos como en el **Teorema V.1**, donde la distribución marginal de cada vector \underline{z}_i de dimensión p , tal que $\underline{z}' = (\underline{z}'_1 \underline{z}'_2 \dots \underline{z}'_n)$, sea $\underline{z}_i \sim EC_p(\underline{\mu}, \frac{\underline{\Sigma}}{\alpha_i}, g)$. Los estimadores puntuales que se obtendrían por el método de Pseudo Máxima Verosimilitud para los parámetros $\underline{\mu}$ y $\underline{\Sigma}$ planteado en la **Proposición V.5**. son los correspondientes estimadores de Máxima Verosimilitud del vector \underline{z} utilizando $\underline{z} = \text{vec}(X')$.

Demostración: Es directa a partir de lo que establece el **Teorema III.1**.

Las dos proposiciones previas establecen una alternativa natural para generalizar el **Teorema V.1** para distribuciones elípticas, sin embargo, debido a que la justificación que llevó al planteamiento de dicho teorema parte del concepto de Log-Verosimilitud Ponderada, es importante establecer la similitud con respecto a dicho concepto que se establece en esta alternativa de generalización. Por ello es importante plantear la comparación entre la función de Log-Verosimilitud Ponderada y la función de Pseudo Máxima Verosimilitud para el caso de distribuciones elípticas.

La función de Log-Verosimilitud Ponderada para el caso de distribuciones elípticas corresponde a:

$$\begin{aligned}
\log[L^*] &= \sum_{i=1}^n \alpha_i \log \left[|\Sigma|^{-\frac{1}{2}} g \left((\underline{x}_i - \underline{\mu})' \Sigma^{-1} (\underline{x}_i - \underline{\mu}) \right) \right] \\
&= \sum_{i=1}^n -\frac{1}{2} \alpha_i \log [|\Sigma|] + \sum_{i=1}^n \log \left[g \left((\underline{x}_i - \underline{\mu})' \Sigma^{-1} (\underline{x}_i - \underline{\mu}) \right) \right]^{\alpha_i} \\
&= -\frac{n}{2} \log [|\Sigma|] + \sum_{i=1}^n \log \left[g \left((\underline{x}_i - \underline{\mu})' \Sigma^{-1} (\underline{x}_i - \underline{\mu}) \right) \right]^{\alpha_i}
\end{aligned}$$

Por otra parte, la función de Pseudo Máxima Verosimilitud para el caso de distribuciones elípticas de acuerdo con la **Proposición V.5** corresponde a:

$$\begin{aligned}
PV &= n^{-1} \sum_{i=1}^n \log \left[\left| \frac{\Sigma}{\alpha_i} \right|^{-\frac{1}{2}} g \left((\underline{x}_i - \underline{\mu})' \left(\frac{\Sigma}{\alpha_i} \right)^{-1} (\underline{x}_i - \underline{\mu}) \right) \right] \\
&= n^{-1} \sum_{i=1}^n -\frac{1}{2} \log [|\Sigma|] + n^{-1} \sum_{i=1}^n \frac{1}{2} \log [\alpha_i] + n^{-1} \sum_{i=1}^n \log \left[g \left((\underline{x}_i - \underline{\mu})' \left(\frac{\Sigma}{\alpha_i} \right)^{-1} (\underline{x}_i - \underline{\mu}) \right) \right] \\
&= n^{-1} \left[-\frac{n}{2} \log [|\Sigma|] + \sum_{i=1}^n \frac{1}{2} \log [\alpha_i] + \sum_{i=1}^n \log \left[g \left(\alpha_i \left((\underline{x}_i - \underline{\mu})' \Sigma^{-1} (\underline{x}_i - \underline{\mu}) \right) \right) \right] \right]
\end{aligned}$$

Para fines de optimización, el valor n^{-1} no es importante, al igual que el segundo sumando. El primer sumando es exactamente igual al correspondiente a la Log-Verosimilitud Ponderada. La diferencia radica en el tercer sumando, en particular en el papel que juega α_i . Para el caso de la distribución Normal Multivariada, ambas expresiones son equivalentes en verosimilitud dado que la función g es exponencial. Para el resto de las distribuciones, es claro que se generaría una aproximación a la Log-Verosimilitud Ponderada al usar la Pseudo Máxima verosimilitud antes presentada, la cual dependerá de la forma funcional de $g(t)$. Evaluar la precisión de la aproximación referida es un problema complejo, que requiere un amplio análisis adicional.

V.2 Estimación Puntual de los Parámetros por el Método de Pseudo Máxima Verosimilitud, en el caso del Análisis de Componentes Principales Comunes en Vectores Dependientes

Para esta sección se parte de la **Proposición V.5**, teniendo siempre en cuenta que la aproximación que plantea dicha proposición, es igualdad para el caso de la distribución Normal Multivariada.

Lema V.1. Sea $\underline{z} \sim EC_{np}(\underline{1} \otimes \underline{\mu}, d \otimes \underline{\Sigma}, g^*)$ con $d = \text{diag}\left(\frac{1}{\alpha_1}, \frac{1}{\alpha_2}, \dots, \frac{1}{\alpha_n}\right)$, para α_i definidos como en el **Teorema V.1**, entonces la verosimilitud de \underline{z} es:

$$L = k |\underline{\Sigma}|^{-n/2} g^*(\text{tr}(\underline{\Sigma}^{-1} G^*))$$

donde

$$k = |d|^{-p/2} \quad \text{y} \quad G^* = \sum_{i=1}^n \alpha_i (\underline{z}_i - \underline{\mu})(\underline{z}_i - \underline{\mu})'$$

para

$$\underline{z}' = (\underline{z}'_1 \underline{z}'_2 \dots \underline{z}'_n)$$

Demostración:

La verosimilitud de \underline{z} es:

$$\begin{aligned} L &= |d \otimes \underline{\Sigma}|^{-1/2} g^* \left[(\underline{z} - \underline{1} \otimes \underline{\mu})(d \otimes \underline{\Sigma})^{-1} (\underline{z} - \underline{1} \otimes \underline{\mu}) \right] \\ &= |d|^{-p/2} |\underline{\Sigma}|^{-n/2} g^* \left(\sum_{i=1}^n \alpha_i (\underline{z}_i - \underline{\mu})' \underline{\Sigma}^{-1} (\underline{z}_i - \underline{\mu}) \right) \\ &= k |\underline{\Sigma}|^{-n/2} g^* \left(\text{tr} \left(\sum_{i=1}^n \alpha_i (\underline{z}_i - \underline{\mu})' \underline{\Sigma}^{-1} (\underline{z}_i - \underline{\mu}) \right) \right) \\ &= k |\underline{\Sigma}|^{-n/2} g^* \left(\sum_{i=1}^n \text{tr} \left(\alpha_i (\underline{z}_i - \underline{\mu})' \underline{\Sigma}^{-1} (\underline{z}_i - \underline{\mu}) \right) \right) \\ &= k |\underline{\Sigma}|^{-n/2} g^* \left(\sum_{i=1}^n \text{tr} \left(\underline{\Sigma}^{-1} \alpha_i (\underline{z}_i - \underline{\mu})(\underline{z}_i - \underline{\mu})' \right) \right) \\ &= k |\underline{\Sigma}|^{-n/2} g^* \left(\text{tr} \left(\sum_{i=1}^n \underline{\Sigma}^{-1} \alpha_i (\underline{z}_i - \underline{\mu})(\underline{z}_i - \underline{\mu})' \right) \right) \\ &= k |\underline{\Sigma}|^{-n/2} g^* \left(\text{tr} \left(\underline{\Sigma}^{-1} \sum_{i=1}^n \alpha_i (\underline{z}_i - \underline{\mu})(\underline{z}_i - \underline{\mu})' \right) \right) \end{aligned}$$

QED

Las **Proposiciones V.5 y V.6** plantean el uso de la verosimilitud antes descrita para generar inferencias considerando los ponderadores de una muestra de diseño complejo, utilizando $\underline{z} = \text{vec}(X')$.

Lema V.2. Sea $\text{vec}(X') \sim EC_{np}(\underline{1} \otimes \underline{\mu}, d \otimes \underline{\Sigma}, g^*)$ con $d = \text{diag}\left(\frac{1}{\alpha_1}, \frac{1}{\alpha_2}, \dots, \frac{1}{\alpha_n}\right)$, para α_i estando definidos como en el **Teorema V.1**, entonces la verosimilitud de $\text{vec}(X')$ es:

$$L = k |\underline{\Sigma}|^{-n/2} g^* \left(\text{tr}(\underline{\Sigma}^{-1} W^*) + \text{tr}(\underline{\Sigma}^{-1} [2\bar{\underline{x}}^* \bar{\underline{x}}^{*'} - n(\bar{\underline{x}}^* \bar{\underline{x}}^{*'}) - 2\underline{\mu} \bar{\underline{x}}^{*'} + n\underline{\mu} \underline{\mu}']) \right)$$

donde

$$k = |d|^{-p/2}$$

$$W^* = \sum_{i=1}^n \alpha_i (\underline{x}_i - \bar{\underline{x}}^*) (\underline{x}_i - \bar{\underline{x}}^*)'$$

$$\bar{\underline{x}}^* = \sum_{i=1}^n \alpha_i \underline{x}_i$$

Demostración:

Partiendo del resultado del **Lema V.1** se desarrollará el término $\text{tr}(\underline{\Sigma}^{-1} G^*)$.

$$\begin{aligned} \text{tr}(\underline{\Sigma}^{-1} G^*) &= \text{tr} \left(\underline{\Sigma}^{-1} \sum_{i=1}^n \alpha_i (\underline{x}_i - \underline{\mu}) (\underline{x}_i - \underline{\mu})' \right) \\ &= \text{tr} \left(\underline{\Sigma}^{-1} \sum_{i=1}^n \alpha_i (\underline{x}_i - \bar{\underline{x}}^* + \bar{\underline{x}}^* - \underline{\mu}) (\underline{x}_i - \bar{\underline{x}}^* + \bar{\underline{x}}^* - \underline{\mu})' \right) \end{aligned}$$

De donde se tiene que:

$$\begin{aligned} \sum_{i=1}^n \alpha_i (\underline{x}_i - \bar{\underline{x}}^* + \bar{\underline{x}}^* - \underline{\mu}) (\underline{x}_i - \bar{\underline{x}}^* + \bar{\underline{x}}^* - \underline{\mu})' &= \sum_{i=1}^n \alpha_i (\underline{x}_i - \bar{\underline{x}}^*) (\underline{x}_i - \bar{\underline{x}}^*)' + \sum_{i=1}^n \alpha_i (\bar{\underline{x}}^* - \underline{\mu}) (\underline{x}_i - \bar{\underline{x}}^*)' \\ &\quad + \sum_{i=1}^n \alpha_i (\underline{x}_i - \bar{\underline{x}}^*) (\bar{\underline{x}}^* - \underline{\mu})' + \sum_{i=1}^n \alpha_i (\bar{\underline{x}}^* - \underline{\mu}) (\bar{\underline{x}}^* - \underline{\mu})' \end{aligned}$$

Este término es igual a:

$$\begin{aligned}
& \sum_{i=1}^n \alpha_i (\underline{x}_i - \bar{x}^*) (\underline{x}_i - \bar{x}^*)' + \sum_{i=1}^n \alpha_i (\bar{x}^* \underline{x}'_i) - \sum_{i=1}^n \alpha_i (\bar{x}^* \bar{x}^{*'}) - \sum_{i=1}^n \alpha_i (\underline{\mu} \underline{x}'_i) + \sum_{i=1}^n \alpha_i (\underline{\mu} \bar{x}^{*'}) \\
& + \sum_{i=1}^n \alpha_i (\underline{x}_i \bar{x}^{*'}) - \sum_{i=1}^n \alpha_i (\bar{x}^* \bar{x}^{*'}) - \sum_{i=1}^n \alpha_i (\underline{x}_i \underline{\mu}') + \sum_{i=1}^n \alpha_i (\bar{x}^* \underline{\mu}') + \sum_{i=1}^n \alpha_i (\bar{x}^* \bar{x}^{*'}) - \sum_{i=1}^n \alpha_i (\bar{x}^* \underline{\mu}') \\
& - \sum_{i=1}^n \alpha_i (\underline{\mu} \bar{x}^{*'}) + \sum_{i=1}^n \alpha_i (\underline{\mu} \underline{\mu}')
\end{aligned}$$

Simplificando términos se obtiene:

$$\begin{aligned}
& \sum_{i=1}^n \alpha_i (\underline{x}_i - \bar{x}^*) (\underline{x}_i - \bar{x}^*)' + 2 \sum_{i=1}^n \alpha_i (\bar{x}^* \underline{x}'_i) - \sum_{i=1}^n \alpha_i (\bar{x}^* \bar{x}^{*'}) - 2 \sum_{i=1}^n \alpha_i (\underline{\mu} \underline{x}'_i) + \sum_{i=1}^n \alpha_i (\underline{\mu} \underline{\mu}') \\
& = \sum_{i=1}^n \alpha_i (\underline{x}_i - \bar{x}^*) (\underline{x}_i - \bar{x}^*)' + 2 \bar{x}^* \sum_{i=1}^n \alpha_i \underline{x}'_i - (\bar{x}^* \bar{x}^{*'}) \sum_{i=1}^n \alpha_i - 2 \underline{\mu} \sum_{i=1}^n \alpha_i \underline{x}'_i + \underline{\mu} \underline{\mu}' \sum_{i=1}^n \alpha_i \\
& = \sum_{i=1}^n \alpha_i (\underline{x}_i - \bar{x}^*) (\underline{x}_i - \bar{x}^*)' + 2 \bar{x}^* \bar{x}^{*'} - n (\bar{x}^* \bar{x}^{*'}) - 2 \underline{\mu} \bar{x}^{*'} + n \underline{\mu} \underline{\mu}'
\end{aligned}$$

De donde finalmente se tiene que:

$$tr(\Sigma^{-1} G^*) = tr \left(\Sigma^{-1} \sum_{i=1}^n \alpha_i (\underline{x}_i - \bar{x}^*) (\underline{x}_i - \bar{x}^*)' \right) + tr \left(\Sigma^{-1} [2 \bar{x}^* \bar{x}^{*'} - n (\bar{x}^* \bar{x}^{*'}) - 2 \underline{\mu} \bar{x}^{*'} + n \underline{\mu} \underline{\mu}'] \right)$$

QED

Nótese que la segunda traza de la verosimilitud, evaluada en $\underline{\mu} = \bar{x}^*$ se hace cero. Estos lemas permiten establecer el siguiente teorema.

Teorema V.2. Sean $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$ vectores que provienen de una población con distribución Elíptica $\underline{x}_i \sim EC_p(\underline{\mu}, \Sigma, g)$, seleccionados por un esquema de diseño muestral complejo. Entonces para obtener una aproximación a los estimadores puntuales que se obtendrían por el método de Pseudo Máxima Verosimilitud para los parámetros β y Λ del modelo de Análisis de Componentes Principales Comunes en Vectores Dependientes, basta con aplicar la Teoría desarrollada para el caso en que la muestra no es de diseño complejo, pero utilizando la matriz W^* en lugar de la matriz W previamente definida.

Demostración:

La verosimilitud obtenida en el **Lema V.2** evaluada en $\underline{\mu} = \bar{x}^*$ que corresponde al estimador de Horvitz-Thompson de $\underline{\mu}$ sería equivalente a la presentada en el **Capítulo IV**, en la ecuación (IV.2) si se sustituyera W por W^* . Esto implica que toda la teoría presentada en el **Capítulo IV** es aplicable si se utiliza la matriz W^* en lugar de la matriz W .

QED

V.3 Inferencia sobre los Parámetros en el caso de muestreos complejos

Cuando se analiza a un conjunto de datos provenientes de una encuesta de diseño muestral complejo, se requiere tener ciertas consideraciones para poder utilizar las técnicas de estimación de varianza descritas en la **Sección IV.4**. A continuación se describen dichas consideraciones para el caso de una encuesta estratificada y multietápica en cada estrato.

Dado que las técnicas descritas en la **Sección IV.4** requieren de la formación de grupos, éstos deberán ser conformados de manera tal que cada uno de ellos sea una submuestra estratificada de la muestra original. Esto significa que cada estrato debe considerarse de manera separada para conformar una parte de los grupos, es decir, que cada estrato aportará una parte (proporcional al tamaño muestral del estrato con respecto a la muestra original) a cada uno de los grupos.

Debido a que en cada estrato se tienen etapas, la conformación de la parte proporcional de cada grupo, obtenida en cada estrato, se tiene que hacer seleccionando de manera separada, y nuevamente de forma proporcional al tamaño de cada unidad primaria de muestreo (UPM), las unidades últimas de muestreo que conforman a cada una de las referidas unidades primarias de muestreo. Esto significa que cada unidad primaria de muestreo aportará una parte proporcional a su tamaño de elementos a cada uno de los subgrupos que se formarán en cada estrato, y que unidos generarán a los grupos requeridos.

Una vez formados los grupos de la manera antes descrita, se pueden utilizar los mismos estimadores que se presentaron en la **Sección IV.4**. Cabe destacar el hecho de que, bajo esta estrategia, no se consideran las unidades inferiores a las primarias de muestreo. Esto se hace bajo la hipótesis de que la variabilidad que aportan las unidades inferiores de muestreo no es considerable, como lo muestran los resultados teóricos que se refieren a la obtención de la varianza de estimadores (Wolter, 1985) y la experiencia empírica.

Las consideraciones descritas previamente, aplican para cualquier estimación de varianza de estimadores que se requiera realizar. En particular es útil en el caso de la estimación de la varianza de los elementos de la estimación de la matriz β obtenida para el modelo de Componentes Principales Comunes en Vectores Dependientes.

También pueden ser utilizadas estas consideraciones si se desea realizar una comparación de las Componentes Principales Comunes entre un grupo (o tiempo) y otro, por ejemplo, mediante la obtención de la varianza estimada de las medias de las Componentes, lo cual permitirá, recurriendo al Teorema Central del límite, realizar pruebas de hipótesis de comparación de las referidas medias.

Capítulo VI. Aplicación del Análisis de Componentes Principales Comunes en Vectores Dependientes Bajo el Supuesto de Distribución Elíptica en Muestreo Aleatorio Simple

VI.1 El Índice de Marginación de México. 1990-2000

En México, en los años noventas, se planteó la necesidad de evaluar numéricamente la condición de marginación de las comunidades del país. El objetivo fué jerarquizarlas en términos de sus necesidades, para contar con un criterio de selección de las comunidades donde implementar programas sociales. De esta necesidad surgió el índice de marginación a nivel municipal y a nivel de localidad, que desde 1990 ha construido el Consejo Nacional de Población (CONAPO) con periodicidad quinquenal.

El mencionado índice fué construido utilizando el Análisis de Componentes Principales aplicado a un conjunto de indicadores del censo de 1990, y seleccionando únicamente a la primera Componente Principal resultante (De la Vega, 1994). Con la necesidad de actualizar el referido índice se recurrió a la aplicación de la misma metodología para los indicadores del conteo de 1995 y posteriormente para el año 2000 se decidió incluso cambiar ciertos indicadores. Todo lo anterior evidenció la imposibilidad de realizar comparaciones inter-temporales de los valores del índice de marginación.

Por ello, en este apartado se presenta la aplicación de un Análisis de Componentes Principales Comunes en Vectores Dependientes, como una alternativa para hacer comparable inter-temporalmente al índice de marginación previamente descrito. Para dicho fin se homologaron los indicadores que se utilizaron para construir los índices de los años 1990 y 2000. Para realizar dicha homologación se buscó, para aquéllos indicadores que no fueran exactamente comparables entre ambos periodos, la posibilidad de reconstruirlos o transformarlos de manera que sí fueran exactamente comparables, utilizando para ello información adicional a partir de otras variables presentes en los censos. Uno de los indicadores no se pudo reconstruir o transformar para ser exactamente comparable para ambos años, por lo cual fué excluído del análisis. Este trabajo de homologación dió lugar al conjunto de indicadores que se presentan en la siguiente tabla.

Tabla VI.1 Indicadores homologados para los años 1990 y 2000

Indicadores:

Porcentaje de población analfabeta de 15 años o más (*analf*)

Porcentaje de viviendas particulares con piso de tierra (*pisotier*)

Porcentaje viviendas particulares sin agua entubada (*sinag*)

Porcentaje de viviendas particulares sin drenaje (*sindr*)

Porcentaje viviendas particulares sin energía eléctrica (*sinel*)

Porcentaje de población ocupada en el sector primario (*secpri*)

Promedio de ocupantes por cuarto en viviendas particulares* (*ocup*)

* El Promedio de ocupantes por cuarto en viviendas particulares se utiliza en el análisis transformado logarítmicamente y reescalado, para que esta variable tenga escalas similares al resto y para reducir su asimetría.

A partir de los indicadores previamente expuestos, y para el nivel de localidad utilizando los datos de 75,006 de ellas (se excluyen localidades confidenciales y con datos faltantes), se muestran por una parte, las ventajas en términos de comparabilidad que implica realizar un Análisis de Componentes Principales Comunes en Vectores Dependientes con respecto a realizar Análisis de Componentes Principales aislados para cada momento en el tiempo. Por otra parte se analiza la evolución de la marginación, utilizando los resultados del mismo Análisis de Componentes Principales Comunes en Vectores Dependientes.

VI.1.1 Comparación entre la construcción de un índice de marginación con un Análisis de Componentes Principales aislados y un Análisis de Componentes Principales Comunes en Vectores Dependientes

Las **Tablas VI.2 y VI.3** presentan los Coeficientes de las Componentes Principales, cuando se realizan para cada uno de los años 1990 y 2000. Existen diferencias en dichos coeficientes, lo cual implica la imposibilidad de realizar comparaciones entre las componentes de los distintos años. La **Tabla VI.4** muestra los coeficientes únicos para aplicarse en ambos años, obtenidos a partir de un Análisis de Componentes Principales Comunes en Vectores Dependientes bajo el supuesto de distribución elíptica. Se puede apreciar cierto patrón de comportamiento de los coeficientes con respecto a los obtenidos a partir del Análisis de Componentes Principales tradicional.

Tabla VI.2 Coeficientes de las Componentes Principales para 1990

Variable	β_1	β_2	β_3	β_4	β_5	β_6	β_7
analf	0.192	0.173	-0.025	-0.007	0.106	-0.123	0.952
pisotier	0.447	0.355	0.011	0.106	0.775	-0.048	-0.246
sinag	0.447	-0.596	-0.615	-0.101	0.015	-0.235	-0.030
sindr	0.223	-0.061	-0.167	0.140	-0.032	0.943	0.088
sinel	0.584	-0.262	0.710	-0.207	-0.208	-0.002	-0.030
secpri	0.202	0.031	0.030	0.919	-0.279	-0.183	-0.032
ocup	0.365	0.644	-0.298	-0.266	-0.516	-0.061	-0.151

Tabla VI.3 Coeficientes de las Componentes Principales para 2000

Variable	β_1	β_2	β_3	β_4	β_5	β_6	β_7
analf	0.163	0.079	0.016	-0.005	0.025	0.120	0.976
pisotier	0.445	0.211	0.110	-0.077	-0.001	0.837	-0.196
sinag	0.440	-0.745	-0.464	0.129	0.136	0.024	-0.012
sindr	0.371	0.054	-0.068	-0.245	-0.861	-0.230	-0.017
sinel	0.419	-0.195	0.799	0.257	0.103	-0.263	-0.036
secpri	0.235	0.038	0.033	-0.853	0.404	-0.225	-0.030
ocup	0.461	0.593	-0.357	0.352	0.255	-0.335	-0.083

**Tabla VI.4 Coeficientes de las Componentes Principales
Comunes en Vectores Dependientes**

Variable	β_1	β_2	β_3	β_4	β_5	β_6	β_7
analf	0.150	0.066	-0.013	0.046	-0.015	-0.134	0.976
pisotier	0.434	0.233	-0.026	0.147	-0.002	-0.833	-0.204
sinag	0.469	-0.639	0.551	-0.135	-0.222	0.029	-0.014
sindr	0.273	0.014	0.172	0.088	0.930	0.156	-0.009
sinel	0.509	-0.255	-0.784	-0.131	-0.023	0.205	-0.038
secpri	0.215	0.097	0.047	0.896	-0.210	0.307	-0.043
ocup	0.435	0.677	0.224	-0.362	-0.205	0.357	-0.046

En la **Tabla VI.5** se pueden apreciar similitudes en los porcentajes de varianza explicada por componente para los Análisis de Componentes Principales tradicionales realizados para cada año. Es destacable el hecho de que la primera componente representa apenas alrededor de la mitad de la varianza explicada. Este hecho hace manifiesta la necesidad de utilizar más de una componente en análisis posteriores derivados de éstos. En particular refleja la necesidad de considerar un índice multivariado de marginación. La **Tabla VI.6** muestra que no existe una diferencia notoria en la distribución de la varianza explicada por componente cuando se utiliza el Análisis de Componentes Principales Comunes en Vectores Dependientes.

**Tabla VI.5 Porcentaje de Varianza Explicada
por los Componentes Principales**

Componente	Individual		Acumulada	
	1990	2000	1990	2000
CP1	51.6%	48.6%	51.6%	48.6%
CP2	14.3%	16.8%	65.9%	65.4%
CP3	12.5%	11.1%	78.4%	76.5%
CP4	8.3%	8.7%	86.7%	85.2%
CP5	6.0%	7.2%	92.7%	92.4%
CP6	4.3%	5.5%	97.0%	97.9%
CP7	3.0%	2.1%	100.0%	100.0%

**Tabla VI.6 Porcentaje de Varianza Explicada
por los Componentes Principales Comunes
en Vectores Dependientes**

Componente	Individual		Acumulada	
	1990	2000	1990	2000
CP1	50.9%	47.8%	50.9%	47.8%
CP2	14.1%	16.7%	65.0%	64.5%
CP3	12.8%	11.3%	77.8%	75.8%
CP4	8.3%	8.8%	86.1%	84.6%
CP5	6.3%	5.5%	92.4%	90.1%
CP6	4.4%	7.7%	96.8%	97.8%
CP7	3.2%	2.1%	100.0%	100.0%

A partir de las **Tablas VI.7 y VI.8**, que presentan las correlaciones entre las Componentes Principales tradicionales y las variables originales, se puede determinar cuáles variables, en qué medida, son representadas en cada componente. Esto es de gran utilidad para un análisis evolutivo pues se puede determinar a partir del estudio de las componentes, la evolución de las variables originales. Sin embargo, como se ha mencionado previamente, estos datos no son comparables por derivarse de distintos análisis.

Tabla VI.7 Correlaciones entre las Componentes Principales para 1990 y las variables de 1990

Variable	CP1	CP2	CP3	CP4	CP5	CP6	CP7
analf	0.61	0.29	-0.04	-0.01	0.11	-0.11	0.72
pisotier	0.80	0.34	0.01	0.08	0.48	-0.02	-0.11
sinag	0.71	-0.50	-0.48	-0.06	0.01	-0.11	-0.01
sindr	0.60	-0.09	-0.22	0.15	-0.03	0.74	0.06
sinel	0.83	-0.20	0.50	-0.12	-0.10	0.00	-0.01
secpri	0.47	0.04	0.03	0.85	-0.22	-0.12	-0.02
ocup	0.65	0.61	-0.26	-0.19	-0.32	-0.03	-0.06

Tabla VI.8 Correlaciones entre las Componentes Principales para 2000 y las variables de 2000

Variable	CP1	CP2	CP3	CP4	CP5	CP6	CP7
analf	0.61	0.18	0.02	-0.02	0.04	0.15	0.75
pisotier	0.81	0.23	0.09	-0.06	0.00	0.51	-0.08
sinag	0.66	-0.66	-0.34	0.08	0.08	0.01	-0.01
sindr	0.72	0.06	-0.07	-0.20	-0.65	-0.15	-0.01
sinel	0.70	-0.19	0.64	0.18	0.07	-0.15	-0.02
secpri	0.51	0.05	0.03	-0.78	0.33	-0.16	-0.01
ocup	0.72	0.54	-0.27	0.23	0.15	-0.17	-0.03

En el caso del Análisis de Componentes Principales Comunes en Vectores Dependientes, se puede apreciar observando las **Tablas VI.9 y VI.10**, que las variables y su grado de representación en cada componente, son prácticamente iguales para ambos años. Esto, aunado a que provienen de un mismo análisis, confirma la comparabilidad de las Componentes Principales Comunes, además de dar muestra nuevamente de la necesidad de utilizar a más de una Componente Principal Común como índice de marginación.

Tabla VI.9 Correlaciones entre las Componentes Principales Comunes en Vectores Dependientes y las variables de 1990

Variable	CP1	CP2	CP3	CP4	CP5	CP6	CP7
analf	0.60	0.19	-0.09	0.02	-0.25	-0.11	0.86
pisotier	0.80	0.22	-0.15	0.10	-0.62	-0.07	0.08
sinag	0.72	-0.57	0.33	-0.09	0.02	-0.13	0.06
sindr	0.62	-0.10	0.12	0.11	0.07	0.67	0.05
sinel	0.80	-0.26	-0.66	-0.03	0.04	-0.06	0.07
secpri	0.47	0.04	-0.05	0.83	0.21	-0.18	0.02
ocup	0.68	0.56	0.11	-0.26	0.06	-0.22	0.19

Tabla VI.10 Correlaciones entre las Componentes Principales Comunes en Vectores Dependientes y las variables de 2000

Variable	CP1	CP2	CP3	CP4	CP5	CP6	CP7
analf	0.60	0.24	0.06	0.09	-0.15	0.12	0.79
pisotier	0.81	0.31	0.05	0.16	-0.52	0.19	-0.04
sinag	0.68	-0.54	0.52	-0.05	0.05	-0.06	0.00
sindr	0.69	0.12	0.23	0.14	0.09	0.82	0.03
sinel	0.73	-0.17	-0.51	-0.02	0.09	0.12	-0.01
secpri	0.50	0.13	0.10	0.84	0.25	-0.05	-0.01
ocup	0.70	0.65	0.25	-0.20	0.19	0.04	0.07

VI.1.2 Análisis Estadístico de los Coeficientes de las Componentes Principales Comunes en Vectores Dependientes para el caso del análisis de la marginación

La estimación asintótica de las varianzas de los elementos de la Matriz $\hat{\beta}$ es:

$1.05E-06$	$3.03E-05$	$1.40E-05$	$2.95E-05$	$3.94E-06$	$4.74E-05$	$1.12E-05$
$3.53E-05$	$2.54E-04$	$7.40E-05$	$1.73E-04$	$1.04E-05$	$1.46E-05$	$4.54E-05$
$9.56E-06$	$1.12E-04$	$1.89E-04$	$4.86E-05$	$2.72E-05$	$5.85E-05$	$1.47E-04$
$2.41E-05$	$2.09E-04$	$6.32E-05$	$1.01E-05$	$8.92E-06$	$2.07E-04$	$6.45E-05$
$1.05E-05$	$1.17E-04$	$4.01E-05$	$6.12E-05$	$1.96E-05$	$4.95E-05$	$2.83E-04$
$2.51E-05$	$4.66E-05$	$1.30E-04$	$2.04E-04$	$1.62E-05$	$2.91E-04$	$1.13E-04$
$1.59E-05$	$1.25E-04$	$2.34E-04$	$8.89E-05$	$2.63E-05$	$8.45E-05$	$5.57E-05$

Se realizan estimaciones de la referida matriz por métodos de remuestreo encontrándose que, para 100 subgrupos generados (elección basada en costos computacionales), en el caso del método de Replicaciones Repetidas se obtiene como estimación:

$1.24E-06$	$3.97E-05$	$6.69E-05$	$4.04E-05$	$6.19E-06$	$2.54E-04$	$1.08E-04$
$3.39E-05$	$4.38E-04$	$2.38E-03$	$2.35E-04$	$1.80E-05$	$6.13E-03$	$2.08E-03$
$6.55E-06$	$2.43E-03$	$1.27E-03$	$5.37E-05$	$4.93E-05$	$2.81E-03$	$6.09E-03$
$2.89E-05$	$2.82E-04$	$6.60E-05$	$8.60E-06$	$1.25E-05$	$6.36E-04$	$1.29E-04$
$9.35E-06$	$1.53E-04$	$3.43E-03$	$6.28E-05$	$4.00E-05$	$1.76E-03$	$1.11E-03$
$2.19E-05$	$4.37E-03$	$3.02E-04$	$2.21E-04$	$2.44E-05$	$1.08E-03$	$7.55E-03$
$1.83E-05$	$6.69E-03$	$4.63E-04$	$1.87E-04$	$4.80E-05$	$1.73E-03$	$4.27E-03$

Por su parte el método de Grupos Aleatorios genera la estimación:

$3.60E-05$	$1.28E-04$	$3.55E-05$	$6.91E-05$	$7.70E-05$	$1.40E-04$	$5.74E-05$
$1.01E-04$	$7.78E-04$	$2.82E-04$	$4.02E-04$	$6.79E-04$	$2.09E-03$	$6.09E-04$
$1.49E-05$	$2.86E-04$	$1.83E-04$	$1.20E-04$	$1.54E-03$	$1.97E-04$	$1.37E-03$
$5.80E-05$	$3.66E-04$	$1.06E-04$	$3.48E-04$	$4.77E-04$	$4.67E-04$	$3.02E-04$
$1.72E-05$	$3.32E-04$	$2.99E-03$	$1.44E-04$	$1.63E-04$	$3.64E-04$	$3.75E-04$
$5.26E-05$	$2.84E-03$	$1.59E-04$	$4.54E-04$	$2.73E-04$	$6.51E-04$	$4.06E-04$
$4.10E-05$	$1.01E-03$	$5.44E-04$	$2.82E-04$	$9.65E-04$	$9.09E-04$	$9.66E-05$

Mientras que el método de Jackknife obtiene como estimación:

$1.49E-06$	$3.91E-05$	$1.57E-05$	$3.91E-05$	$7.45E-06$	$5.24E-05$	$1.38E-05$
$4.18E-05$	$4.21E-04$	$9.48E-05$	$2.07E-04$	$2.46E-05$	$2.24E-05$	$6.92E-05$
$8.22E-06$	$1.20E-04$	$4.83E-04$	$4.78E-05$	$5.67E-05$	$6.20E-05$	$3.17E-04$
$3.04E-05$	$2.62E-04$	$6.39E-05$	$8.72E-06$	$1.35E-05$	$2.57E-04$	$9.56E-05$
$1.19E-05$	$1.20E-04$	$8.28E-05$	$7.39E-05$	$5.06E-05$	$6.52E-05$	$4.70E-04$
$3.22E-05$	$8.88E-05$	$1.37E-04$	$2.41E-04$	$2.46E-05$	$5.38E-04$	$1.14E-04$
$1.81E-05$	$2.30E-04$	$3.32E-04$	$2.01E-04$	$5.14E-05$	$1.58E-04$	$1.82E-04$

Se puede observar similitud entre los resultados, lo cual es de esperarse dado el gran tamaño de muestra que se analizó.

VI.1.3 Análisis evolutivo de la marginación en México de 1990 a 2000

Una de las principales ventajas al utilizar el Análisis de Componentes Principales Comunes en Vectores Dependientes es que las comparaciones inter temporales son válidas de manera directa. Por ello, si se requiere analizar la evolución de la marginación en México, se puede analizar el cambio en los valores de las Componentes Principales Comunes. Para interpretar claramente las diferencias entre estas Componentes Principales Comunes, es conveniente presentar las diferencias en las variables entre 1990 y 2000, que se presentan a continuación.

Tabla VI.11 Medias de las variables en los años 1990 y 2000

Variable	Media en 1990	Media en 2000
analf	28.72	23.21
pisotier	54.61	43.54
sinag	66.78	50.70
sindr	86.99	71.03
sinel	53.77	27.65
secpri	73.94	64.82
ocup	72.75	83.85

La tabla anterior muestra el promedio de cada variable, y debido a que la unidad de análisis es, en este caso, la localidad, estos promedios no pueden ser interpretados de ningún modo como promedios a nivel nacional, ya que no se está considerando el tamaño (distinto) de las localidades. Ahora bien, en el contexto que se está analizando, el tamaño de una localidad se ignora porque lo que interesa es el cambio a nivel de localidad, no a nivel de hogar ni de individuo.

Debido a que todos los indicadores están expresados de forma tal que un mayor valor implica mayor marginación, entonces una disminución puede interpretarse como una mejora. Se puede apreciar que, a excepción de los ocupantes por cuarto (que se registra transformada), el resto de las variables tuvieron una mejora entre 1990 y 2000. En particular se observa una notoria disminución en el valor promedio en la variable que refleja el porcentaje de viviendas sin energía eléctrica. Con respecto a la variable relativa a los ocupantes por cuarto, sería interesante analizar si el cambio que en primera instancia parece negativo, se debe a una mejora producto de la disminución en la mortalidad, en la migración o a un cambio de la pirámide poblacional reflejando un conjunto mayor de personas económicamente activas.

A continuación se presentan las diferencias entre las medias de las Componentes Principales Comunes para 1990 y 2000. La significancia que se incluye corresponde al valor p de la prueba de comparación de medias para datos pareados, y se observa que en todos los casos se concluye que la diferencia es estadísticamente significativa.

Tabla VI.12 Diferencia de medias entre Componentes Principales Comunes

Variable	Diferencia de medias	Significancia
CP1	28.10	0.000
CP2	-20.41	0.000
CP3	-11.27	0.000
CP4	9.86	0.000
CP5	-2.81	0.000
CP6	11.00	0.000
CP7	1.90	0.000

La interpretación de estas diferencias no es directa como en el caso de las variables originales en el sentido de afirmar que una diferencia positiva refleja una mejoría. En este caso hay que recurrir tanto a los coeficientes de las Componentes Principales Comunes, como a las Correlaciones entre Componentes y variables originales. Por ejemplo, para la primera componente se observa en la **Tabla VI.4** que todos los coeficientes son positivos, conservándose con ello el sentido de la interpretación de que una disminución en el valor de la componente implica una mejora. Así pues al observarse una notoria diferencia positiva (la mayor de todas) en esta componente, se puede concluir que las variables asociadas con esa componente, que de acuerdo con las **Tablas VI.9 y VI.10** son la mayoría, mejoraron de 1990 a 2000.

Al interpretar la segunda componente se puede apreciar en las **Tablas VI.9 y VI.10** que las variables relacionadas fuertemente con ella son el porcentaje de ocupantes y el porcentaje de viviendas sin agua. Se destaca de la **Tabla VI.4** que el coeficiente para la primera variable referida es positivo mientras que el de la segunda es negativo. En el sentido original de la interpretación esto es un problema ya que se diría por un lado que una disminución en la componente refleja mejora (de acuerdo con la variable porcentaje de ocupantes) y por otro se plantearía lo contrario para la variable correspondiente al porcentaje de viviendas sin agua.

La tercera componente refleja de acuerdo con las **Tablas VI.9 y VI.10** los cambios en la variable porcentaje de viviendas sin electricidad, y debido a que el coeficiente de dicha variable

en esa tercera componente es negativo, entonces la diferencia negativa entre los valores de dicha componente de 1990 a 2000, puede interpretarse en este caso como una mejora, en menor cuantía que la mejora reflejada por la primera componente.

Como puede apreciarse, para realizar una interpretación adecuada de los resultados obtenidos se deben considerar tanto los coeficientes de las Componentes Principales Comunes como las correlaciones entre las componentes y las variables, además de las diferencias entre las mismas Componentes Principales Comunes. La interpretación del resto de las diferencias entre componentes de distintos periodos se realiza de manera análoga a las previamente mostradas, considerando los puntos señalados. Un caso interesante se observa en la séptima componente que se asocia completamente al concepto de analfabetismo. No se presenta aquí el resto de la interpretación debido a que la **Tabla VI.6** muestra que las tres primeras componentes explican la mayor parte de la varianza, (alrededor de un 80 por ciento).

Capítulo VII. Aplicación del Análisis de Componentes Principales Comunes en Vectores Dependientes Bajo el Supuesto de Distribución Elíptica en Diseños Muestrales Complejos

VII.1 Encuesta Nacional de los Niveles de Vida de los Hogares, México 2002-2005

La Encuesta Nacional de los Niveles de Vida de los Hogares (ENNViH), se diseñó con múltiples propósitos entre los que se encuentra la evaluación del impacto de programas de apoyo a la pobreza, de mejora educativa y de salud, de disminución de la criminalidad y la violencia entre otros. Para cumplir su propósito, la encuesta es de carácter longitudinal, lo cual implica que se da seguimiento a los individuos, hogares y comunidades encuestados de manera periódica. Actualmente se dispone de la información correspondiente a la ENNViH-1 cuyo levantamiento se realizó en 2002, pero aún no se publican los datos correspondientes a la información de la ENNViH-2 cuyo levantamiento se realizó en 2005, por esta razón, en este trabajo se utilizarán datos simulados.

La ENNViH tiene un diseño muestral complejo, bietápico en algunos segmentos y trietápico en otros, incluyendo estratificación, por lo cual para ser correctamente utilizada requiere de la consideración del diseño muestral complejo, particularmente de sus ponderadores. El tamaño de la muestra es de 8,440 hogares, que representan a una población de aproximadamente 25,000,000 de hogares que habitaban el país en 2002.

La información que recaba es multitemática, recopilándose datos a nivel individual, de hogar y de comunidad. Para efectos de presentar un ejemplo con datos reales que a su vez sea didáctico se han seleccionado solamente unas secciones de la encuesta para ser analizadas. La selección se basa en los resultados de estudios económicos que han conducido a clasificar las condicionantes del Nivel de Vida de un Hogar de acuerdo a los siguientes dos aspectos: a) los activos y b) el capital humano.

En lo que respecta a los activos, se plantea la hipótesis de que éstos se pueden agrupar a su vez en activos permanentes, como las características de la vivienda, y activos temporales, como los enseres domésticos, los ingresos y los ahorros. Con respecto al capital humano se puede mencionar que el nivel de escolaridad es la medida más común. Así pues, las variables que se seleccionaron para ejemplificar la técnica estadística se relacionan con estas agrupaciones de indicadores. Es importante señalar de manera adicional que, dada la intención de evaluar los cambios en el Nivel de Vida, se descartaron aquéllas variables que difícilmente cambiarían en el lapso de tres años, como por ejemplo la escolaridad del jefe del hogar.

Tabla VII.1 Variables de la ENNViH

Variables:

Disponibilidad de excusado en la vivienda (*excusado*)
Disponibilidad de un piso de material firme en la vivienda (*pisofirme*)
Ingreso per cápita del hogar* (*ingreso*)
Porcentaje de personas entre 18 y 65 años (*pers18.65*)
Porcentaje de niños entre 5 y 18 años que asisten a la escuela (*asiste5_18*)
Disponibilidad de enseres eléctricos (*enseres*)
Disponibilidad de aparatos electrónicos (*electrónicos*)
Disponibilidad de ahorros (*ahorros*)

* El Ingreso per cápita del hogar se utiliza en el análisis transformado logarítmicamente, para reducir la asimetría de la variable original.

La disponibilidad de excusado en la vivienda (bajo la premisa de que este indicador refleja de manera intrínseca la disponibilidad de agua y de drenaje) y la disponibilidad de un piso de material firme en la vivienda se seleccionaron como reflejo de las características que difícilmente se pierden una vez adquiridas, considerándose por ello activos permanentes.

El ingreso per cápita del hogar es una variable que por sí sola es indicativo de la condición de un hogar.

El porcentaje de personas entre 18 y 65 años, es decir en edad de sumarse a la población económicamente activa, y el porcentaje de niños entre 5 y 18 años que asisten a la escuela se tomaron como medidas del capital humano por el potencial de generación de ingreso que tienen actualmente o tendrán a futuro los individuos.

La disponibilidad de enseres eléctricos (lavadora y refrigerador) y la disponibilidad de aparatos electrónicos (televisión y estéreo) se eligieron como indicadores de los enseres en general.

Finalmente se escogió la disponibilidad de ahorros como reflejo del activo disponible para contingencias.

Debe observarse que este conjunto de variables incluye algunas que son categóricas y de hecho dicotómicas. Naturalmente, en este caso el supuesto distribucional de elipticidad o normalidad no puede sostenerse. En la práctica, sin embargo, es frecuente el empleo de variables categóricas y por ello, este ejemplo se ha incluido en el presente trabajo con el objetivo de examinar los resultados que genera el Análisis de Componentes Principales Comunes en Vectores Dependientes cuando se aplica a datos de este tipo con un propósito exploratorio. Evidentemente, los procedimientos inferenciales que se fundamentan en el supuesto distribucional no se aplican en este caso.

Se generaron los resultados para los Análisis de Componentes Principales Comunes en Vectores Dependientes con estas variables, primero para el caso en que se ignora el hecho de que existe un diseño muestral complejo, y luego considerando dicho diseño, con el objetivo de comparar los resultados en ambos casos. A continuación se presenta dicha comparación.

VII.1.1 Comparación de los resultados de un Análisis de Componentes Principales Comunes en Vectores Dependientes cuando se ignora el diseño muestral complejo y cuando se considera el diseño muestral complejo

Dado que los datos de la segunda ronda de la encuesta ENNViH no se encuentran disponibles aún, se decidió simularlos a partir de las siguientes consideraciones:

1) Para simular el efecto de la tendencia de una mejora generalizada esperada, se asigna a cada hogar una probabilidad aleatorizada de mejorar en cada variable de manera proporcional a su ingreso.

2) Para simular el efecto de programas de apoyo a los menos favorecidos se asignó una probabilidad de mejora en las variables que reflejan la disponibilidad de piso firme en la vivienda (pisofirme) y enseres eléctricos (enseres), que es inversamente proporcional al ingreso per cápita.

Utilizando los datos simulados para el año 2005 y los datos reales del año 2002, se ejemplifica el Análisis de Componentes Principales Comunes en Vectores Dependientes, utilizando los datos de cada uno de los dos años como un grupo. En particular se realiza una comparación de los resultados obtenidos cuando se ignora el diseño muestral complejo, y cuando éste se toma en cuenta.

A continuación se presentan los coeficientes de las Componentes Principales Comunes en Vectores Dependientes, para los datos descritos. La existencia de diferencias significativas entre ambas tablas refleja el error que se comente al no considerar el diseño muestral complejo. Esto conduce a la necesidad de determinar si las diferencias que se observan en las tablas, son estadísticamente significativas.

Tabla VII.2 Componentes Principales Comunes en Vectores Dependientes sin considerar el diseño muestral complejo

Variables	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
excusado	0.977	-0.039	-0.154	0.000	-0.002	0.043	-0.005	-0.134
pisofirme	0.135	-0.005	0.721	0.000	-0.001	-0.677	0.000	-0.064
ingreso	-0.003	-0.001	0.000	0.950	-0.090	0.001	-0.299	-0.006
pers18_65	0.004	0.002	-0.003	0.312	0.209	-0.002	0.927	-0.005
asiste5_18	0.001	0.005	0.000	0.021	0.974	-0.001	-0.227	0.001
enseres	0.063	0.005	0.672	0.000	0.001	0.734	0.003	-0.075
electrónicos	0.147	0.006	0.077	0.007	-0.001	0.018	0.002	0.986
ahorros	0.037	0.999	-0.007	0.000	-0.005	-0.005	-0.001	-0.011

Tabla VII.3 Componentes Principales Comunes en Vectores Dependientes considerando el diseño muestral complejo

Variables	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
excusado	0.963	-0.028	-0.171	0.002	-0.007	-0.105	-0.007	-0.176
pisofirme	0.177	-0.002	0.560	0.002	0.000	0.807	0.001	-0.058
ingreso	-0.006	-0.004	0.001	0.949	-0.084	-0.002	-0.303	-0.007
pers18_65	0.006	0.004	-0.005	0.314	0.182	0.000	0.932	-0.004
asiste5_18	0.006	0.008	-0.002	0.023	0.980	0.000	-0.199	0.002
enseres	0.059	0.006	0.805	-0.001	0.002	-0.579	0.004	-0.116
electrónicos	0.191	0.003	0.098	0.008	-0.003	-0.039	0.001	0.976
ahorros	0.027	0.999	-0.008	0.002	-0.009	0.002	-0.004	-0.008

La sexta componente ha presentado un cambio de signo en todo el vector, debido a que la definición de las columnas es única salvo por permutaciones y multiplicaciones por menos uno. Una vez que se ha cambiado el signo de dicho coeficiente en alguna de las dos tablas, se puede hacer la valoración de la magnitud de las variaciones, que resultan ser las más notorias. El hecho de que a simple vista, las diferencias en la sexta componente sean las más grandes, no se puede interpretar como que estadísticamente sean las más significativas, pues es necesario para ello medir la variabilidad asociada con las mismas. Sin embargo, como ya se mencionó, las variables dicotómicas no permite utilizar los resultados asociados a la inferencia de las componentes.

Para poder determinar las causas de las diferencias, se puede partir de un análisis preliminar sobre las variables que determinan cada componente. Mediante la observación directa de la magnitud de los coeficientes asignados a cada variable dentro de cada componente, se puede apreciar que, por ejemplo, el coeficiente de la variable excusado, en la primera componente, es notoriamente más grande que el resto, lo cual indica que esta variable determina en cierta medida a dicha componente. Las variables pisofirme y electrónicos tienen coeficientes de valor similar entre sí, y aunque mucho menores que el de la variable excusado, también son mucho mayores que los del resto de las variables. Esto implica que estas variables también pueden estar determinando, en distinta medida, a la primera componente. Esto plantea la necesidad de realizar un análisis más profundo que permita evaluar de manera estadística si estos coeficientes son significativos en la determinación de la primera componente, y con ello poder determinar las causas de las diferencias en dicha componente al asociarlas con las diferencias en las variables que la determinan.

En el caso particular de la sexta componente y partiendo del análisis preliminar, se puede observar que los mayores coeficientes corresponden a las variables pers18_65 y enseres. Al ser esa componente la que más cambia a simple vista cuando se considera el diseño muestral complejo, una primera hipótesis que se podría plantear es que las variables pers18_65 y enseres son las que más se afectan con la ponderación de los datos. Esto refleja que un análisis preliminar puede conducir a plantear hipótesis, que en ningún caso deben considerarse concluyentes.

Aún en los casos en que este análisis preliminar de los coeficientes de las componentes principales asociados a las variables originales, pueda parecer muy claro, como es el caso de la segunda componente, en la que la variable ahorros es la que determina prácticamente de manera total a dicha componente, se hace manifiesta la conveniencia de hacer un análisis inferencial para validar las relaciones entre las Componentes obtenidas y las variables originales. Una alternativa para validar dichas relaciones de manera descriptiva es recurrir

al cálculo de las combinaciones lineales que generan las componentes a nivel de hogar para cada uno de los tiempos, y obtener las correlaciones entre las variables originales y las componentes. A continuación se presentan dichas correlaciones, resaltando el hecho de que al calcular correlaciones entre variables continuas y dicotómicas se obtiene una medida de asociación.

Tabla VII.4 Correlación entre Variables y Componentes Principales Comunes en Vectores Dependientes (2002) sin considerar el diseño muestral complejo

Variable	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8
excusado	0.988	0.129	0.256	0.164	0.106	-0.007	0.071	-0.071
pisofirme	0.492	0.091	0.820	0.122	0.078	-0.532	0.058	0.051
ingreso	0.149	0.131	0.115	0.963	-0.061	0.026	-0.082	0.075
pers18_65	0.162	0.066	0.104	0.507	0.203	-0.004	0.912	0.058
asiste5_18	0.099	0.089	0.051	0.008	0.978	-0.021	-0.221	-0.039
enseres	0.384	0.112	0.794	0.132	0.046	0.606	0.046	0.297
electrónicos	0.332	0.096	0.457	0.152	0.005	0.224	0.072	0.948
ahorros	0.211	0.999	0.110	0.144	0.081	0.016	0.000	0.026

Tabla VII.5 Correlación entre Variables y Componentes Principales Comunes en Vectores Dependientes (2002) considerando el diseño muestral complejo

Variable	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8
excusado	0.981	0.144	0.289	0.169	0.106	0.030	0.088	-0.114
pisofirme	0.565	0.115	0.708	0.136	0.074	0.671	0.074	0.029
ingreso	0.149	0.128	0.132	0.962	-0.068	-0.015	-0.084	0.073
pers18_65	0.181	0.083	0.110	0.510	0.197	0.023	0.919	0.055
asiste5_18	0.104	0.067	0.037	0.011	0.983	0.028	-0.165	-0.054
enseres	0.439	0.117	0.889	0.147	0.036	-0.436	0.050	0.264
electrónicos	0.396	0.099	0.534	0.166	-0.008	-0.169	0.082	0.918
ahorros	0.206	0.999	0.119	0.147	0.051	0.011	0.018	0.017

Tabla VII.6 Correlación entre Variables y Componentes Principales Comunes en Vectores Dependientes (2005) sin considerar el diseño muestral complejo

Variable	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8
excusado	0.991	0.125	0.077	0.183	0.096	0.008	0.077	-0.082
pisofirme	0.373	0.058	0.783	0.067	0.059	-0.629	0.045	0.036
ingreso	0.166	0.138	0.038	0.966	-0.056	0.019	-0.079	0.104
pers18_65	0.173	0.070	0.057	0.513	0.196	0.001	0.908	0.082
asiste5_18	0.088	0.082	0.031	0.010	0.978	-0.016	-0.228	-0.026
enseres	0.260	0.071	0.710	0.070	0.034	0.634	0.039	0.183
electrónicos	0.298	0.086	0.341	0.181	0.014	0.123	0.081	0.953
ahorros	0.205	0.999	0.049	0.151	0.073	0.007	0.002	0.020

Tabla VII.7 Correlación entre Variables y Componentes Principales Comunes en Vectores Dependientes (2005) considerando el diseño muestral complejo

Variable	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8
excusado	0.985	0.138	0.101	0.183	0.092	-0.028	0.083	-0.108
pisofirme	0.443	0.081	0.632	0.084	0.056	0.737	0.061	0.026
ingreso	0.163	0.138	0.064	0.966	-0.062	-0.028	-0.080	0.102
pers18_65	0.183	0.089	0.080	0.518	0.190	-0.003	0.914	0.082
asiste5_18	0.094	0.063	0.018	0.014	0.982	0.014	-0.172	-0.035
enseres	0.295	0.079	0.833	0.093	0.030	-0.502	0.055	0.179
electrónicos	0.367	0.088	0.417	0.192	0.004	-0.120	0.092	0.929
ahorros	0.197	0.999	0.063	0.157	0.046	0.001	0.020	0.016

Para fines de interpretación se han resaltado las componentes con una correlación mayor a 0.5 en al menos una de las tablas. Esto no implica que las correlaciones marcadas sean estadísticamente significativas. Se deben realizar pruebas de hipótesis para determinar la significancia estadística de las correlaciones, para lo cual es necesario cuantificar la variabilidad asociada. Se resalta en este punto el hecho de que la variabilidad asociada a un estimador, cuando se trata de una encuesta de diseño muestral complejo, depende de la estructura de dicho diseño, por lo cual para realizar un análisis adecuado se requiere tener en cuenta dicha estructura.

De las correlaciones se reiteran algunas de las conclusiones que se obtuvieron en el análisis preliminar, pero ahora sustentados estadísticamente. Se aprecia que la primera componente tiene asociada fuertemente a la variable excusado, y que tiene una asociación media con la variable pisofirme. Se puede observar por un lado el hecho de que la asociación con la variable electrónicos es menor que la asociación con la variable pisofirme, aún cuando sus coeficientes eran muy similares, asimismo se aprecia que existe un grado de asociación con la variable enseres similar y en ocasiones mayor al grado de asociación correspondiente a la variable electrónicos, aún cuando en los coeficientes esto no se reflejó.

Lo observado en el análisis preliminar con respecto a la segunda y a la sexta componentes se valida, al ser los ahorros la variable con mayor asociación con respecto a la segunda componente, y al ser por su parte tanto los enseres como el pisofirme, las variables con mayor asociación con respecto a la sexta componente.

Un análisis integral de la información proporcionada por las tablas anteriores conduce a concluir que las primeras cinco componentes se asocian en general a grupos distintos de variables. La tabla siguiente refleja dichas asociaciones. Se destaca el hecho de que las asociaciones que se muestran en la siguiente tabla, aplican no importando si se considera o no el efecto de diseño, pero esto no puede generalizarse de ninguna manera.

Tabla VII.8 Variables mayormente asociadas a cada Componente

Componente	Variable(s) Asociada(s)		
CP1	excusado	pisofirme	
CP2	ahorros		
CP3	enseres	electrónicos	pisofirme
CP4	ingreso	pers18_65	
CP5	asiste5_18		
CP6	pisofirme	enseres	
CP7	pers18_65		
CP8	electrónicos		

De la tabla anterior se puede apreciar que la hipótesis original con respecto a la posibilidad de obtener un conjunto de nuevas variables que concentren la información de las variables originales por grupos bien definidos puede ser viable. Se observa que los activos permanentes se asocian con la primera componente, que los ahorros se asocian con la segunda, que los activos temporales se asocian con la tercera (a excepción del pisofirme), que los ingresos en conjunto con la estructura sociodemográfica del hogar se asocian con la cuarta componente, y que el capital humano en formación se asocia con la quinta componente. A partir de ahí se comienzan a repetir las asociaciones. Si bien el pisofirme es la única variable que rompe con la hipótesis, cabe recordar que fué una de las variables simuladas, por lo cual este efecto podría diluirse en los datos reales. Esto implica que este conjunto de cinco componentes podría ser útil para realizar una comparación por grupos de variables de la evolución de los individuos en el tiempo.

Se presenta a continuación la variabilidad explicada por cada componente, lo cual permite determinar la proporción de la información original que cada componente principal obtenida refleja. Se puede observar de dicha información que tanto en el caso de no considerar el diseño muestral complejo como en el caso en que se considera el mismo, la varianza global explicada por las cinco primeras componentes ronda el 80 por ciento, con lo cual realizar un análisis con esas cinco componentes implicaría reflejar el 80 por ciento de la información, lo cual en este contexto puede ser muy útil.

Tabla VII.9 Porcentaje de Varianza Explicada por los Componentes Principales Comunes en Vectores Dependientes sin considerar el diseño muestral complejo

Componente	Individual		Acumulada	
	2002	2005	2002	2005
CP1	28.0%	28.4%	28.0%	28.4%
CP2	17.3%	18.7%	45.3%	47.1%
CP3	12.1%	8.8%	57.4%	55.9%
CP4	12.0%	13.4%	69.4%	69.3%
CP5	9.2%	9.7%	78.6%	79.0%
CP6	7.9%	8.4%	86.4%	87.4%
CP7	7.0%	7.0%	93.5%	94.4%
CP8	6.5%	5.6%	100.0%	100.0%

**Tabla VII.10 Porcentaje de Varianza Explicada
por los Componentes Principales Comunes
en Vectores Dependientes
considerando el diseño muestral complejo**

Componente	Individual		Acumulada	
	2002	2005	2002	2005
CP1	25.4%	25.3%	25.4%	25.3%
CP2	19.4%	20.8%	44.8%	46.1%
CP3	12.8%	14.2%	57.5%	60.3%
CP4	11.7%	8.6%	69.3%	68.9%
CP5	9.2%	9.7%	78.5%	78.6%
CP6	8.2%	8.7%	86.7%	87.3%
CP7	6.7%	6.8%	93.4%	94.1%
CP8	6.6%	5.9%	100.0%	100.0%

Aunque no radicales, sí se observan diferencias notorias en la distribución de la explicación de la variabilidad entre componentes cuando se considera el diseño muestral complejo, y este es un reflejo de los cambios en los pesos (ponderadores) de las observaciones, que a su vez influyen en el peso de las variables en cada una de las componentes.

VII.1.2 Análisis evolutivo de las condiciones de vida entre 2002 y 2005 en México

De acuerdo con lo expuesto, analizar la evolución de las condiciones de vida de los hogares en México entre los años 2002 y 2005 utilizando el Análisis de Componentes Principales Comunes en Vectores Dependientes, puede conducir a realizar análisis evolutivos por grupos de variables simultáneamente, en vez de realizarlos variable por variable. Para poder apreciar la ventaja de dicha posibilidad, a continuación se presenta el análisis variable por variable, y posteriormente se realiza utilizando las componentes principales obtenidas previamente. Adicionalmente para continuar con la comparación, se presentan los resultados cuando no se considera el diseño muestral complejo y se comparan con los obtenidos cuando si se considera el mismo.

Si bien ya se ha mencionado que hay problemas para realizar inferencias en este ejemplo por tener variables dicotómicas, hacer inferencias sobre las medias de las componentes resultantes, una vez calculadas las combinaciones lineales, puede realizarse por medio de pruebas no paramétricas como la Wilcoxon.

**Tabla VII.10 Medias de las variables en los años 2002 y 2005
sin considerar el diseño muestral complejo**

Variable	Media en 2002	Media en 2005	Diferencia	Significancia
excusado	71.7%	73.7%	-0.020	0.000
pisofirme	87.3%	92.8%	-0.055	0.000
ingreso	940.4	1395.5	-455.1	0.000
pers18_65	57.8%	58.8%	-0.010	0.000
asiste5_18	88.4%	88.8%	-0.004	0.000
enseres	89.2%	94.6%	-0.053	0.000
electrónicos	92.2%	93.6%	-0.014	0.000
ahorros	16.6%	17.1%	-0.005	0.000

**Tabla VII.11 Medias de las variables en los años 2002 y 2005
considerando el diseño muestral complejo**

Variable	Media en 2002	Media en 2005	Diferencia	Significancia
excusado	78.8%	80.3%	-0.016	0.000
pisofirme	88.9%	93.4%	-0.045	0.000
ingreso	1169.8	1694.0	-524.2	0.000
pers18_65	59.1%	60.1%	-0.010	0.000
asiste5_18	89.2%	89.5%	-0.003	0.000
enseres	89.2%	94.4%	-0.052	0.000
electrónicos	92.5%	93.7%	-0.012	0.000
ahorros	17.9%	18.3%	-0.004	0.000

En las tablas previas se puede apreciar que las diferencias más notorias entre periodos se dan en las variables pisofirme y enseres, lo cual era de esperarse. Al comparar las tablas se observa que existen variables que aún considerando el diseño muestral complejo, no tienen

grandes modificaciones en promedio, mientras que existen otras que reflejan un mayor cambio en promedio. Esto puede implicar que en ciertas ocasiones el considerar el diseño puede no ser tan importante en este punto cuando se analizan por separado aquellas variables que permanecen sin cambio, sin embargo para efectos de un análisis multivariado como el que se desea realizar, es importante considerar el diseño muestral complejo pues al menos una de las variables incluidas en el análisis difiere de manera considerable.

Tabla VII.12 Diferencia de medias entre Componentes Principales Comunes sin considerar el diseño muestral complejo

Variable	Diferencia de medias	Significancia
CP1	-0.032	0.000
CP2	-0.004	0.000
CP3	-0.073	0.001
CP4	-0.019	0.000
CP5	-0.004	0.000
CP6	-0.003	0.000
CP7	-0.004	0.000
CP8	-0.003	0.000

Tabla VII.13 Diferencia de medias entre Componentes Principales Comunes considerando el diseño muestral complejo

Variable	Diferencia de medias	Significancia
CP1	-0.029	0.000
CP2	-0.004	0.480
CP3	-0.065	0.642
CP4	-0.020	0.000
CP5	-0.004	0.000
CP6	-0.004	0.000
CP7	-0.004	0.000
CP8	0.000	0.000

Las tablas previas muestran variaciones significativas en la mayoría de las componentes, aunque de distinta magnitud. En este caso si el interés se centra en analizar las variaciones de esas primeras componentes por lo antes expuesto, se llega a las mismas conclusiones sobre la

prueba de diferencia de medias considerando el diseño con respecto a cuando no se considera. Sin embargo la significancia estadística tanto de la segunda como de la tercera componentes, muestran que es posible encontrar diferencias significativas cuando no se considera el diseño muestral complejo, que ya no lo son cuando se considera dicho diseño, y viceversa.

Es importante destacar que aún diferencias pequeñas como en el caso de la quinta componente (cuya diferencia es notoriamente menor a la obtenida en la primera y tercera componentes) resultan significativas en este caso, debido en gran parte a que se está analizando una base de datos de gran tamaño. Estos resultados refleja la utilidad en construir las componentes pues permiten de manera agrupada, evaluar el comportamiento de conjuntos de variables. Lo cual es muy trascendente cuando se intenta realizar un análisis comparativo de un atributo que está medido por medio de varias variables, ya que de este modo se evalúa su evolución de manera conjunta.

Comentarios finales y conclusiones

Previo a la realización del presente proyecto de investigación se conocían los resultados que permitían llevar a cabo un Análisis de Componentes Principales Comunes en Vectores Dependientes bajo el supuesto de Normalidad, aplicable a encuestas obtenidas bajo muestreo aleatorio irrestricto.

Adicionalmente se conocían diversos resultados concernientes a alternativas distribucionales a la Normal Multivariada, como el caso de las distribuciones elípticas, así como herramientas para la aplicación de métodos estadísticos a datos provenientes de encuestas de diseño muestral complejo.

Derivado del gran potencial de aplicación que tiene el Análisis de Componentes Principales Comunes en Vectores Dependientes, particularmente en el área de estudios sociales donde preferentemente se dispone de encuestas de diseño muestral complejo, el presente trabajo se avocó a aportar los elementos necesarios para llevar a cabo un Análisis de Componentes Principales Comunes en Vectores Dependientes bajo un supuesto distribucional más fácil de validar que el de Normalidad, y que pudiera aplicarse a encuestas obtenidas bajo diseño muestral complejo. Los principales resultados obtenidos fueron:

A) Los estimadores de máxima verosimilitud para los parámetros del modelo de Componentes Principales Comunes en Vectores Dependientes bajo el supuesto de distribución Elíptica.

B) La prueba de cociente de verosimilitudes generalizado para definir la validez del modelo de Componentes Principales Comunes en Vectores Dependientes bajo el supuesto de distribución Elíptica.

C) La distribución asintótica de los estimadores de máxima verosimilitud para los parámetros del modelo de Componentes Principales Comunes en Vectores Dependientes bajo el supuesto de distribución Elíptica.

D) La demostración de la equivalencia en la obtención de los estimadores de máxima verosimilitud para los parámetros del modelo de Componentes Principales Comunes en Vectores Dependientes bajo los supuestos de Normalidad y Elípticidad, y con ello la posibilidad de utilizar las mismas herramientas computacionales para ambos supuestos distribucionales.

E) La demostración de que el estimador de máxima verosimilitud para el parámetro β del modelo de Componentes Principales Comunes en Vectores Dependientes es igual bajo los supuestos de distribución Normal y Elíptica.

F) Una propuesta de pseudoverosimilitud para incorporar el diseño muestral complejo a la estimación de parámetros de un modelo.

G) La demostración de la equivalencia entre la aplicación de la pseudo verosimilitud propuesta y el método mediante el uso de una distribución nominal con ajuste en la varianza para el caso Normal, y la posibilidad de extender el resultado al caso elíptico.

H) La demostración de la equivalencia en la obtención de los estimadores de máxima verosimilitud para los parámetros del modelo de Componentes Principales Comunes en Vectores Dependientes bajo el supuesto de distribución Elíptica aplicado a datos provenientes de una encuesta de diseño muestral complejo utilizando la matriz de varianzas muestral ponderada, con respecto a la obtención de dichos estimadores utilizando la matriz de varianzas muestral en el caso de tener muestreo aleatorio irrestricto. Esto implicando nuevamente la posibilidad de usar las mismas herramientas computacionales para ambos tipos de diseños muestrales.

En términos generales, este conjunto de resultados hace posible llevar a cabo un Análisis de Componentes Principales Comunes en Vectores Dependientes bajo el supuesto de elíptica, aplicable a encuestas obtenidas bajo diseño muestral complejo.

A partir de estos resultados se obtienen las siguientes conclusiones y comentarios:

1. El supuesto de distribuciones elípticas flexibiliza el supuesto de normalidad, puesto que implica la búsqueda de un modelo distribucional para un conjunto multivariado de datos, dentro de una familia mucho más extensa y diversa de distribuciones que la Normal Multivariada, que además, está incluida en dicha familia. Esto quiere decir que se tiene mayor probabilidad de encontrar el modelo distribucional adecuado para el referido conjunto multivariado de datos.

En la actualidad, y debido a lo anterior, se han retomado los estudios sobre la familia de distribuciones elípticas en diversos aspectos estadísticos. Algunos desarrollos destacados relacionados con estadística inferencial son el de los estimadores máximo verosímiles y la distribución tanto teórica como asintótica de los mismos (Fang, et. al., (1990); Steyn, (1996); Lindsey, (1999)).

Por el lado del análisis de bondad de ajuste se tienen diversas propuestas metodológicas para determinar la elipticidad de un conjunto de datos y con ello poder validar el supuesto distribucional (Beran, (1979); Manzotti, et. al. (2002); Huffer, et. al. (2007)).

Adicionalmente se han extendido las propuestas en el campo de la simulación de datos, permitiendo generar muestras de distribuciones elípticas por medio de herramientas que se pueden implementar computacionalmente, utilizando elementos como la representación de Cambanis y metodologías de simulación como el método de Johnson-Ramberg, o por medio de simulación de Monte-Carlo a partir de las distribuciones condicionales, e incluso mediante el uso de distribuciones mezcla (Devroye, (1986); Haro-López, (1998)). Esto ha permitido analizar de manera controlada las ventajas del uso de este tipo de distribuciones en la práctica.

Si bien en este proyecto se utilizaron muchos de los desarrollos descritos, aún queda mucho por hacer en este renglón, por ejemplo evaluar alternativas de simulación y compararlas, así como aplicar las propuestas del análisis de bondad de ajuste y evaluar su viabilidad y eficiencia (mediante simulación) para validar el supuesto de elipticidad. Estos desarrollos se dejan abiertos como líneas de investigación a futuro.

2. La posibilidad de utilizar distribuciones elípticas en diversos análisis que tradicionalmente parten de un supuesto de normalidad, genera la inquietud sobre la posibilidad de que los desarrollos teóricos de los referidos análisis se compliquen hasta el punto en que tal vez no se puedan encontrar soluciones aplicables prácticamente. Al analizar los desarrollos teóricos del modelo de Análisis de Componentes Principales Comunes bajo el supuesto de distribuciones elípticas, presentado en este proyecto, se puede observar que el modelo puede llevarse a la práctica mediante la generalización de la teoría correspondiente al supuesto de normalidad, si bien con ajustes que tienen su grado de complejidad, no imposibles de solventar. Esto puede llevar a la conclusión de que será viable implementar la teoría para otro tipo de modelos como el Análisis Factorial, o en general cualquier otro método multivariado que parte del supuesto de normalidad multivariada de los datos.

De hecho, a partir de los resultados obtenidos para el modelo de Análisis de Componentes Principales Comunes bajo el supuesto de distribuciones elípticas se puede apreciar que incluso algunos resultados son invariantes con respecto al supuesto de normalidad (como en el caso de la estimación de los coeficientes de las Componentes Principales Comunes), mientras que otros resultados se desvían ligeramente al ser modificados solo por constantes. Sólo en el caso de las distribuciones asintóticas de los estimadores, se presentan diferencias notorias, en algunos casos con alto grado de dificultad para poder ser llevadas a la práctica como en el caso de las correspondientes distribuciones del parámetro Λ .

De este modo se puede plantear como proyectos para desarrollar en ésta línea de investigación, los desarrollos teóricos de otros análisis multivariados bajo el supuesto de distribución elíptica, así como el análisis con mayor profundidad de la distribución del parámetro Λ dentro de este mismo Análisis de Componentes Principales Comunes en Vectores Dependientes.

3. Debido a la complejidad reflejada en la obtención directa de la varianza de los estimadores de los modelos, se planteó en este proyecto la posibilidad de explorar la estimación de dicha varianza por medio de métodos como Replicaciones Balanceadas, Grupos aleatorios y Jackknife como alternativas no paramétricas de relativa facilidad de implementación (Rao, (1981); Efron, (1982); Shao, et. al., (1995); Lunnenborg, (2000)). De su implementación en varios ejemplos se observa que la determinación del número de subconjuntos utilizados para realizar los cálculos es un elemento fundamental para la obtención de estimaciones precisas. Adicionalmente se pudo apreciar que en general el método más consistente es el de Jackknife al permitir obtener estimaciones adecuadas tanto para muestras pequeñas como para muestras grandes, aunque bajo una adecuada implementación y mediante la detección del número de subconjuntos adecuado, los tres métodos generan estimaciones similares.

Debido a la complejidad en la obtención de la varianza de estimadores, estas técnicas se presentan como una alternativa viable, bajo ciertas condiciones. En este rubro cabe señalar que, debido a que la determinación del número óptimo de subconjuntos es muy importante, una línea de investigación que surge de este punto es la exploración de metodologías para la determinación de dicho número óptimo de subconjuntos óptimo, particularmente para el caso de estimación de varianzas para parámetros vectoriales o matriciales. Una propuesta consiste en explorar la homogeneidad generada entre los grupos conformados, y determinar el número de subconjuntos a partir de la maximización de dicha homogeneidad. Otra cuestión que surge de este rubro es la correspondiente a la comparación de matrices de varianzas, para poder evaluar el desempeño de los métodos de estimación, para lo cual se plantea la necesidad de explorar medidas de disimilaridad entre matrices.

4. La necesidad que surge de la práctica, derivada de la optimización de recursos, de realizar encuestas con diseños muestrales complejos, conduce a su vez a la necesidad de adecuar las metodologías de análisis estadístico para poder contemplar el efecto que produce el referido diseño muestral complejo. Si bien ya existen esfuerzos en este sentido, se han planteado casi de manera exclusiva en el renglón univariado (Cochran, (1963); Kish, et. al. (1974); Skinner, et. al. (1989)), con algunas excepciones que han buscado analizar el análisis multivariado (Koch, et. al. (1975); Skinner, et. al. (1986); Binder, (1992); Graubard, (1994)) y en pocos casos son propuestas de carácter general. El método de Pseudo Máxima verosimilitud, y en particular el de Log verosimilitud ponderada bajo el supuesto

de normalidad, se proponen como una alternativa viable para realizar análisis multivariados a datos provenientes de encuestas de diseño complejo. Debido a que la Log Verosimilitud Ponderada tiene sentido al proponerse como una forma de ponderar la información que cada individuo de la muestra aporta (en el diseño muestral complejo) hace de este método una alternativa de fácil interpretación. Adicionalmente, la posibilidad de generalizar el método para distribuciones elípticas, hace que su aplicabilidad se incremente.

Una vez implementado el método en el Análisis de Componentes Principales Comunes en Vectores Dependientes bajo el supuesto de distribución elíptica cuando se dispone de una muestra proveniente de un diseño muestral complejo, se puede apreciar que la implementación práctica es viable dado que la estimación se reduce a aplicar las metodologías obtenidas para muestreo aleatorio simple, pero aplicadas a las matrices de varianzas y covarianzas ponderadas. Esto tiene gran sentido, pues es el equivalente multivariado a la estimación de la forma Horvitz-Thompson. Es claro sin embargo que la inferencia sobre los estimadores no es directa. Pero debido a lo descrito en el punto anterior, se puede recurrir a métodos de estimación de varianza como el de Jackknife, Replicaciones Balanceadas o Grupos Aleatorios, debidamente ajustados para contemplar el diseño muestral complejo. En este sentido se han desarrollado ya alternativas de probada eficiencia (Wolter, et. al., (1985); Sarndal, et. al. (1992)).

De este punto surge la posibilidad de extender la línea de investigación hacia la posibilidad de utilizar el método de Log verosimilitud ponderada en otros análisis multivariados, así como la posibilidad de explorar otros métodos alternativos, tanto de estimación como de inferencia (obtención de varianzas) que contemplen el diseño muestral complejo.

5. La implementación computacional de todos los métodos descritos es viable gracias a las herramientas disponibles en la actualidad. Esto es un reflejo del avance de la tecnología que permite llevar a la práctica análisis que antes se consideraban imposibles de utilizar. En particular estos avances conducen a la necesidad de replantear la posibilidad de llevar a la práctica análisis que en algún momento fueron desarrollados de manera teórica y que nunca se consideraron viables de utilización por la complejidad en la obtención de los cálculos. Adicionalmente a estos aspectos, se debe considerar que la tecnología sigue avanzando, de manera que hoy en día se puede disponer de herramientas como el cómputo en paralelo que permite la generación de resultados (principalmente en lo que concierne a simulaciones) con una gran reducción en tiempo computacional, y que han llevado incluso a replantear las consideraciones que se realizan sobre la eficiencia de algoritmos que con esta herramienta pueden considerarse mejores que los que actualmente se encuentran en uso, como por ejemplo los algoritmos iterativos (Wilkinson, et. al. (1999)).

Estas consideraciones llevan a reforzar la necesidad de realizar estudios de carácter interdisciplinario, ya que la necesidad de implementar algoritmos de cómputo de alta complejidad requieren de análisis de tipo numérico que permitan, para poder asegurar que los resultados sean utilizables para generar conclusiones, determinar la convergencia, eficiencia y sensibilidad de los referidos algoritmos, conceptos ampliamente estudiados en el campo del referido análisis numérico (Golub, et. al. (1989)).

La necesidad de colaborar interdisciplinariamente se muestra también a través del desarrollo del presente proyecto en el sentido de utilizar herramientas matemáticas para facilitar los cálculos teóricos, como por ejemplo el recurrir a las herramientas del Álgebra matricial. Esto se ha reflejado en la generación de bibliografía estrechamente interrelacionada entre la Estadística y el Álgebra Matricial (Harville, (1997); Schott, (2005)).

6. En lo relacionado con el modelo de Análisis de Componentes Principales Comunes en Vectores Dependientes, en cuestión de aplicabilidad, es destacable el esfuerzo desarrollado por diversos investigadores para mostrar la gran flexibilidad que el modelo tiene (Flury, (1988); Klingenberg, et. al. (1996); Neuenschwander, et. al. (2000)). En particular del presente proyecto se puede concluir que la aplicabilidad del modelo se puede extender al análisis de encuestas de tipo longitudinal de carácter social, por sus características de comparabilidad intertemporal, permitiendo de esta forma proponerlo como una metodología de evaluación de la evolución de características a lo largo del tiempo.

Las aplicaciones mostradas en el ámbito del análisis de la evolución de la marginación y de la evolución de las condiciones de vida muestran las grandes posibilidades de aplicación del método en el entorno de análisis sociales, destacándose aspectos muy importantes como la posibilidad de evaluar cambios inter espaciales e inter temporales, así como la aplicabilidad a encuestas de diseño muestral complejo, permitiéndolo obtener conclusiones estadísticamente válidas, y que a su vez permiten generar tanto planeación como evaluación de proyectos. Una de las aplicaciones que se tienen es la posibilidad de construir y analizar índices multidimensionales que permitan extender la visión que se tiene en muchas áreas de estudio que se limitan a analizar índices unidimensionales, con la gran ventaja de poder realizar análisis de varias variables de manera simultánea.

Bibliografía

- [1] Agresti, A. (1990). *Categorical Data Analysis*. John Wiley & Sons. University of Florida.
- [2] Anderson, T. W., (1988). *An Introduction to Multivariate Statistical Analysis*. Wiley.
- [3] Beaghen, M. J. (1997). *Canonical Variate Analysis and Related Methods with Longitudinal Data* Tesis.
- [4] Bebbington, A. C. y Smith, T. M. F. (1977). *The Effect of Survey Design on Multivariate Analysis* en *The Analysis of Survey Data*, vol. 2. John Wiley.
- [5] Beran, R. (1979). *Testing for Ellipsoidal Symmetry of a Multivariate Density*. The Annals of Statistics, vol. 7, No. 1, pp. 150-162.
- [6] Binder, D. A. (1992). *Fitting Cox's Proportional Hazards Models From Survey Data*. Cambridge University Press.
- [7] Bishop, Y. M., Fienberg, S. E. y Holland, P. W. (1975). *Discrete Multivariate Analysis*. MIT Press.
- [8] Boente, G., Pires, A. M. y Rodrigues, I. M. (2002). *Influence Functions and Outlier Detection Under the Common Principal Components Model: A Robust Approach*. Biometrika, Vol. 89-4, pp. 861-875.
- [9] Boente, G., Pires, A. M. y Rodrigues, I. M. (200?). *General Projection-pursuit Estimators for the Common Principal Components Model : Influence Functions and Monte Carlo Study*. Journal of Multivariate Analysis, en imprenta.
- [10] Branco, D. y Dey, D. K. (2001). *A General Class of Multivariate Skew-Elliptical Distributions*. Journal of Multivariate Analysis, Vol. 79, pp. 99-113.
- [11] Cochran, W. G. (1963). *Sampling Techniques*. John Wiley & Sons.
- [12] Conner, J. K., Franks, R. y Stewart, C. (2003). *Espression of Additive Genetic Variances and Covariances for Wild Radish Floral Traits: Comparison Between Field and Greenhouse Environments*. Evolution, Vol. 57, No. 3, pp. 487-495.
- [13] Chmielewski, M. A. (1981). *Elliptically Symmetric Distributions: A Review and Bibliography*. International Statistical Review, vol. 49, pp. 67-74.
- [14] De la Vega, S. (1994). *Construcción de un Índice de Marginación. Condiciones de Población y Vivienda de los Municipios de la República Mexicana* Tesis de Maestría.
- [15] Deville, J. C. y Särndal (1996). *Calibration Estimators in Survey Sampling*. Journal of the American Statistical Association, vol. 87, pp. 376-382.
- [16] Devlin, S. J., Gnanadesikan, R. y Kettering, J. R. (1981). *Robust Estimation of Dispersion Matrices and Principal Components*. Journal of the American Statistical Association, vol. 76, No. 374, pp. 354-362.

- [17] Devroye, L. (1986). *Non-Uniform Random Variate Generation*. Springer-Verlag.
- [18] Draper, N. R. y Smith, H., (1998). *Applied Regression Analysis, 3rd. ed.* Wiley Interscience.
- [19] Dunteman, G. H., (1989). *Principal Components Analysis*. Sage Publications.
- [20] Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*. Capital City Press.
- [21] Efron, B. y Hinkley D. V. (1978). *Assessing the Accuracy of the Maximum Likelihood Estimator: Observed versus Expected Fisher Information*. *Biometrika*, vol. 65, no. 3, pp. 457-487.
- [22] Falk, M. (2002). *The Sample Covariance is Not Efficient for Elliptical Distributions*. *Journal of Multivariate Analysis*, vol. 80, pp. 358-377.
- [23] Falk, M. (1998). *A Note on the Comedian for Elliptical Distributions*. *Journal of Multivariate Analysis*, vol. 67, pp. 306-317.
- [24] Fang, K. y Li, R. (1999). *Bayesian Statistical Inference on Elliptical Matrix Distributions*. *Journal of Multivariate Analysis*, vol. 70, pp. 66-85.
- [25] Fang, K. y Anderson, T. W. (1990a). *On the Theory of Multivariate Elliptically Contoured Distributions and Their Applications*. *Statistical Inference in Elliptically Contoured and Related Distributions*. Allerton Press Inc.
- [26] Fang, K. y Anderson, T. W. (1990b). *Inference in Multivariate Elliptically Contoured Distributions Based on Maximum Likelihood*. *Statistical Inference in Elliptically Contoured and Related Distributions*. Allerton Press Inc.
- [27] Fang, K., Kotz, S. y Ng K. (1990). *Symmetric Multivariate and Related Distributions*. Chapman and Hall.
- [28] Fengler, M. R., Hardle, W. K. y Villa, C. (2001). *The Dynamics of Implied Volatilities: A Common Principal Components Approach*. Working paper.
- [29] Fienberg, S. E. (1979). *The Analysis of Cross-Classified Categorical Data*. The Massachusetts Institute of Technology. MIT Press.
- [30] Flury, B. (1986). *Asymptotic Theory for Common Principal Components Analysis*. *The Annals of Statistics*, vol. 14, no. 2, pp. 418-430.
- [31] Flury, B. (1987). *Two Generalizations of the Common Principal Component Model*. *Biometrika*, vol. 14, no. 1, pp. 59-69.
- [32] Flury, B. (1988). *Common Principal Components and Related Multivariate Models*. John Wiley & Sons.
- [33] Flury, B. y Constantine, G. (1984). *The F-G Algorithm*. *Applied Statistics*, vol. 34, no. 2, pp. 177-183.
- [34] Gilbert, J y Gilbert, L. (1995). *Linear Algebra and Matrix Theory*. Academic Press.

- [35] Golub, G. H. y Van Loan, C. F. (1989). *Matrix Computations, 2nd. ed.* The John Hopkins University Press.
- [36] Graubard, B. I. y Korn, E. L. (1994). *Regression Analysis With Clustered Data* Statistics in Medicine, vol. 13, pp. 509-522.
- [37] Graybill, F. A. (1954). *On Quadratic Estimates of Variance Components* Documento de trabajo no. J-2171 de la Estación de experimentos agrícolas de Iowa, pp. 367-372.
- [38] Greenacre, M. J. (1984). *Theory and Applications of Correspondence Analysis.* Academic Press INC.
- [39] Gu, H. y Fung, W. K. (2001). *Influence Diagnostics in Common Principal Components Analysis.* Journal of Multivariate Analysis, vol. 79, pp. 275-294.
- [40] Hair, J. F., Anderson, R. E., Tatham, R. L. y Black, W. C. (1995). *Multivariate Data Analysis.* Prentice Hall. New Jersey.
- [41] Haro-López, R. (1998). *Data Adaptive Bayesian Analysis Using Distributional Mixtures.* Ph.D. Thesis, Imperial College of Science, Technology and Medicine, London, U.K.
- [42] Hashorva, E. (2007). *Conditional Limiting Distribution of Type III Elliptical Random Vectors.* Journal of Multivariate Analysis, vol. 98, pp. 282-294.
- [43] Huffer, F. W. y Park, C. (2007). *A Test for Elliptical Symmetry.* Journal of Multivariate Analysis, vol. 98, pp. 256-281.
- [44] Harville, A. (1997). *Matrix Algebra from a Statistician's Perspective.* Springer-Verlag.
- [45] Henderson, H. V. y Searle, S. R. (1979). *Vec and Vech Operators for Matrices, with Some Uses in Jacobians and Multivariate Statistics.* The Canadian Journal of Statistics vol. 7, no. 1, pp. 65-81.
- [46] Houle, D., Mezey, J. y Galpern, P. (2002). *Interpretation of the Results of Common Principal Components Analyses.* Evolution, Vol. 56, No. 3, pp. 433-440.
- [47] Huberty, C. J. (1994). *Applied Discriminant Analysis.* John Wiley & Sons. New York.
- [48] Jobson, J. D. (1992). *Applied Multivariate Data Analysis. Volume I: Regression and Experimental Design.* Springer-Verlag.
- [49] Jobson, J. D. (1992). *Applied Multivariate Data Analysis. Volume II: Categorical and Multivariate Methods.* Springer-Verlag.
- [50] Johnson, R. A. y Wichern, D. W. (1998). *Applied Multivariate Statistical Analysis.* Prentice Hall.
- [51] Johnson, M. E. (1987). *Multivariate statistical Simulation.* John Wiley & Sons.
- [52] Johnson, N. L., Kotz, S. y Balakrishnan N. (1994). *Continuous Univariate Distributions, vol 1., 2nd. ed.* John Wiley & Sons.

- [53] Jolliffe, I. T. (2002). *Principal Component Analysis, 2nd. ed.* Springer.
- [54] Kish, L. (1965). *Survey Sampling.* Wiley.
- [55] Kish, L. y Frankel, M. R. (1974). *Inferences From Complex Samples.* Journal of the Royal Statistical Society, Serie B, vol. 36, pp. 1-37.
- [56] Kleinbaum, D. (1992) *Logistic Regression. A Self-Learning Text.* Springer Verlag.
- [57] Klingenberg, C. P., Neuenschwander, B. E. y Flury, B. D. (1996) *Ontogeny and Individual Variation: Analysis of Patterned Covariance Matrices with Common Principal Components.* Systematic Biology, Vol. 45, no. 2, pp. 135-150.
- [58] Knuth, D. E. (1984) *The Texbook.* Addison Wesley Publishing Company.
- [59] Koch, G. G., Freeman, D. H. y Freeman, J. L. (1975). *Strategies in the Multivariate Analysis of Data from Complex Surveys.* International Statistical Review, Vol. 43, No. 1, pp. 59-78.
- [60] Koch, G. G. y Lemeshow, S. (1972). *An Application of Multivariate Analysis to Complex Sample Survey Data.* Journal of the American Statistical Association, Vol. 67, No. 340, pp. 780-782.
- [61] Krzanowski, W. J. (1972). *Orthogonal Components for Grouped Data: Review and Applications.* Statistics in Transition, Vol. 5, No. 5, pp. 759-777.
- [62] Kuba, L. (2006). *Componentes Principales Comunes. Un Método para la Comparación de Matrices de Varianza y Covarianza* Tesis.
- [63] Kubokawa, T. y Srivastava, M. S. (1999). *Robust Improvement in Estimation of a Covariance Matrix in an Elliptically Contoured Distribution.* The Annals of Statistics, vol. 27, No. 2, pp. 600-609.
- [64] LaMotte, L. R. (1973). *Quadratic Estimation of Variance Components* Biometrics, vol.29, pp. 311-330.
- [65] Lebart, L., Morineau, A. y Warwick K. M. (1977). *Multivariate Descriptive Statistical Analysis. Correspondence Analysis and Related Techniques for Large Matrices.* John Wiley & Sons.
- [66] Lee, P. M. (1989). *Bayesian Statistics: An Introduction.* Oxford University Press.
- [67] Lindsey, J. K. (1999). *Multivariate Elliptically Contoured Distributions for Repeated Measurements.* Biometrics, vol. 55, pp. 1277-1280.
- [68] Lunneborg, C. E. (2000). *Data Analysis by Resampling: Concepts and Applications.* Duxbury Press.
- [69] Magnus, J. R. y Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics.* Wiley.
- [70] Manzotti, A., Perez, F. J. y Quiroz, A. J. (2002). *A Statistic for Testing the Null Hypothesis of Elliptical Symmetry.* Journal of Multivariate Analysis, vol. 81, pp. 274-285.

- [71] Mezey, J. G. y Houle, D. (2003). *Comparing G Matrices: Are Common Principal Components Informative?* Genetics, Vol. 165, pp. 411-425.
- [72] Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley.
- [73] Nathan, G. y Holt, D. (1980). *The Effect of Survey Design on Regression Analysis*. Journal of the Royal Statistical Society, Serie B, Vol. 42, No. 3, pp. 377-386.
- [74] Neuenschwander, B. E. (1991). *Common Principal Components for Dependent Random Vectors*. Ph. D. Thesis, Univ. Bern, Bern, Switzerland.
- [75] Neuenschwander, B. E. y Flury, B. D. (1994). *Simultaneous Diagonalization Algorithms with Applications in Multivariate Statistics*. Approximation and computation: a festschrift in honor of Walter Gautschi (R. Zahar, ed.). Birkhauser, Boston.
- [76] Neuenschwander, B. E. y Flury, B. D. (1995). *Principal Component Models for Patterned Covariance Matrices, with Applications to Canonical Correlation Analysis of Several Sets of Variables*. Descriptive multivariate analysis (W. J. Krzanowski, ed.). Oxford University Press
- [77] Neuenschwander, B. E. y Flury, B. D. (2000). *Common Principal Components for Dependent Random Vectors*. Journal of Multivariate Analysis, Vol. 75, pp. 163-183.
- [78] Ortega, V., Angeles, A., Medina, J., Landassuri, V. y Figueroa, J. (2006). *Computing Similarity of Square Matrices by Eigenconjugation* Proceedings of the Fifth Mexican International Conference on Artificial Intelligence, IEEE.
- [79] Pervaiz, M. K. y Skinner, C. J. (1992). *Multivariate Analysis of Clustered Survey Data: Testing Equality of Covariance Matrices* Australian Journal of Statistics, Vol. 34(2), pp. 295-306.
- [80] Pervaiz, M. K. y Skinner, C. J., (1990). *A Monte Carlo Comparison of Elliptical-Theory and Other Tests for Equality of Covariance* Australian Journal of Statistics, vol. 32(1), pp. 71-86.
- [81] Pigliucci, M., (2000). *Common Principal Components Analyses* Pigliucci Lab.
- [82] Press, W. H., Flannery, B., Teukolsky, S. A. y Vetterling, W. T. (1988). *Numerical Recipes in C*. Cambridge University Press.
- [83] Rao, J N. K y Scott, A. J. (1981). *The Analysis of Categorical Data From Complex Sample Surveys: Chi-Squared Tests For Goodness of Fit and Independence in Two-Way Tables*. Journal of the American Statistical Association, Vol. 76, pp. 221-230.
- [84] Rao, J N. K y Wu, C. F. J. (1988). *Resampling Inference With Complex Survey Data*. Journal of the American Statistical Association, Vol. 83, pp. 231-241.
- [85] Robert, C. P. (2001). *The Bayesian Choice*. Springer-Verlag.
- [86] Rubalcava, L. y Teruel, G. (2004). *The Mexican Family Life Survey Project (MxFLS): Study Design and Baseline Results*. Documento de trabajo CIDE y UIA.
- [87] Särndal, C. E., Swensson, B. y Wretman, J. (1992). *Model Assisted Survey Sampling*. Springer.

- [88] Schey, H. M. (1997). *Div, Grad, Curl, and All That: an Informal Text on Vector Calculus*. W. W. Norton & Company.
- [89] Schott, J. R. (1991). *Some Tests for Common Principal Component Subspaces in Several Groups*. *Biometrika*, Vol. 78, No. 4, pp. 771-777.
- [90] Schott, J. R. (1991). *Partial Common Principal Component Subspaces*. *Biometrika* Vol. 86, No. 4, pp. 899-908.
- [91] Schott, J. R. (2005). *Matrix Analysis for Statistics*. John Wiley & Sons.
- [92] Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons.
- [93] Shao, J. y Dongsheng, T. (1995). *The Jackknife and Bootstrap*. Springer-Verlag
- [94] Skinner, C. J. y De Toledo Vieira, M. (2005). *Design Effects In The Analysis of Longitudinal Survey Data*. Documento de trabajo M05/13 del Southhampton Statistical Sciences Research Institute.
- [95] Skinner, C. J., Holmes, D. J. y Smith, T. M. F. (1986). *The Effect of Sample Design on Principal Component Analysis*. *Journal of the American Statistical Association*, Vol. 81, No. 395, pp. 789-798.
- [96] Skinner, C. J., Holt, D. y Smith, T. M. F. eds. (1989). *Analysis of Complex Surveys*. Wiley.
- [97] Steyn, H. S. (1996). *The Distribution of the Covariance Matrix for a Subset of Elliptical Distributions with Extension to Two Kurtosis Parameters*. *Journal of Multivariate Analysis*, vol. 58, pp. 96-106.
- [98] Turkington, D. A. (2002). *Matrix Calculus & Zero-One matrices*. Cambridge University Press.
- [99] Wilkinson, B. y Allen, M. (1999). *Parallel Programming*. Prentice-Hall.
- [100] Wolter, K. M. (1985). *Introduction to Variance Estimation*. Springer-Verlag.
- [101] Zirkel, G. y Berlinger, E. (1994). *Understanding Fortran -77 & 90-*. PWS Publishing Company.

Referencias Adicionales

- [1] Bartlett, M. S. (1935). *Contingency Table Interactions*. Journal of the Royal Statistical Society, suppl. 2, pp. 248-252.
- [2] Fisher, R. A. (1936). *The Use of Multiple Measurements in Taxonomic Problems*. Annals of Eugenics.
- [3] Fisher, R. A. (1940). *The Precision of Discriminant Functions*. Annals of Eugenics, vol. 10, pp. 422-429.
- [4] Galton, F. (1885). *Regression Towards Mediocrity in Hereditary Stature*. Journal of the Antropological Institute, vol. 15, pp. 246-263.
- [5] Hirschfeld, H. O. (1935). *A Connection Between Correlation and Contingency*. Cambridge Philosophical Soc. Proc. (Math. Proc.), vol. 31, pp. 520-524.
- [6] Hotelling, H. (1933). *Analysis of a Complex of Statistical Variables into Principal Components*. J. Educ. Psychol., vol. 24, pp. 417-441 y 498-520.
- [7] Hotelling, H. (1936). *Relations Between Two Sets of Variates*. Biometrika, vol. 28, pp. 321-377.
- [8] Pearson, K. (1901). *On Lines and Planes of Closest Fit to Systems of Points in Space*. Philosophical Magazine, vol. 2 (sexta serie), pp. 559-572.
- [9] Spearman, C. (1904). *"General-Intelligence", Objectively Determined and Measured*. American Journal of Psychology, vol. 15, pp. 201-293.
- [10] Yule, G. U. (1900). *On the Association of Attributes in Statistics: with Illustration from the Material of the Childhood Society*. Philos. Trans. Roy. Soc., ser. A., vol. 194, pp. 257-319.

Apéndice: Programa que implementa el algoritmo FG+

```

C
SUBROUTINE FGPALG(A, B, IP, K, IPK, IFAULT, NF)
USE MSIMSL
C      PROGRAMA EN FORTRAN PARA EL ALGORITMO FG+ ORTOGONAL
C      A=MATRIZ DE VARIANZAS COVARIANZAS PARTICIONADA DE IPK X IPK (INPUT)
C      B=MATRIZ DE COMPONENTES PRINCIPALES COMUNES IP X IP (OUTPUT)
C      IP=NUMERO DE VARIABLES (INPUT)
C      K=NUMERO DE GRUPOS (INPUT)
C      IPK= IP * K (INPUT)
C      IFAULT=CODIGO DE RESULTADO (0-SIN ERROR; 1 K-INVALIDO;
C              2-IP INVALIDO; 4-NO CONVERGIO;
C              -I(I>0)-MATRIZ A NO SIMETRICA O NO POSITIVA DEFINIDA
C      NF=NUMERO DE ITERACIONES DEL ALGORITMO F+
C
REAL A(IPK, IPK), B(IP, IP), S(IPK, IPK)
REAL ZERO, ONE, P8, P6, G1, G2
DATA ZERO /0.0/, ONE /1.0/, P8 /0.8/, P6 /0.6/
C
C      CHEQUEO DE LOS PARAMETROS
C
IFAULT = 1
IF (K .LT. 1) RETURN
IFAULT =2
IF (IP .LT. 2) RETURN
DO 4 M = 2, IPK
    IFAULT = -1
    JEND = M - 1
    DO 4 J = 1, JEND
        IF (A(M, J) .NE. A(J, M)) RETURN
4    CONTINUE
IFAULT = 0
C
C      INICIALIZA LA MATRIZ B COMO LA MATRIZ IDENTIDAD
C
DO 5 M = 1, IP
    DO 5 J = 1, IP
        B(M,J) = ZERO
        IF (M .EQ. J) B(M, J) = ONE
5    CONTINUE
C
C      ALMACENA LA MATRIZ A EN S PARA SU USO POSTERIOR
C
DO 7 M = 1, IPK
    DO 7 J = 1, IPK
7        S(M, J) = A(M, J)
C
C      DESARROLLO DEL ALGORITMO F+
C
CALL FPALG(A, B, IP, K, IPK, IFAULT, NF)
IF (IFAULT .NE. 0 .OR. NF .GT. 1) RETURN
C
C      SI EL ALGORITMO F TERMINA LUEGO DE UNA SOLA ITERACION,

```

```

C          SE PRUEBA CON OTRA APROXIMACION INICIAL A LA MATRIZ B
C
DO 8 M = 1, IPK
  DO 8 J = 1, IPK
8      A(M, J) = S(M, J)
  IP1 = IP - 1
  DO 6 M = 1, IP1
    DO 6 J = 1, IP
      G1 = P8 * B(M, J) + P6 * B(M + 1, J)
      G2 = P8 * B(M + 1, J) - P6 * B(M, J)
      B(M, J) = G1
      B(M + 1, J) = G2
6  CONTINUE
  CALL FPALG(A, B, IP, K, IPK, IFAULT, NF)
  RETURN
  END
C
C
C  SUBROUTINE FPALG(A, B, IP, K, IPK, IFAULT, NF)
  REAL AUX(IPK, IPK), B(IP, IP), BOLD(IP, IP), H(IP, 2),
&    Q(2, 2), TX(IPK, IPK), T(2*K, 2*K), A(IPK, IPK)
  REAL ZERO
  DATA ZERO /0.0/, EPSF /1.E-8/, MAXF /100/
  NF = 0
  IFAULT = 0
  IP1 = IP - 1
  CONTINUE
4  C
  C          PASO F1+ (ACTUALIZA LA MATRIZ B)
  C
  NF = NF + 1
  DO 5 M = 1, IP
    DO 5 J = 1, IP
5      BOLD(M, J) = B(M, J)
  C
  C          PASO F2+ (INICIO DE LAS ITERACIONES DEL ALGORITMO F+)
  C
  DO 8 M = 1, IP1
    JSTART=M+1
    DO 8 N = JSTART, IP
  C
  C          PASO F21+ (GENERA LA MATRIZ H)
  C
    DO 6 I = 1, IP
      H(I, 1) = B(I, M)
      H(I, 2) = B(I, N)
6  CONTINUE
  C
  C          PASO F22+ (GENERA LA MATRIZ T)
  C
    CALL KRONI(K, H, IP, 2, AUX)
    DO 12 Z = 1, IPK
      DO 12 W = 1, IPK
        TX(Z, W) = A(Z, W)
12  CONTINUE
    KK=2*K

```

```

                CALL MULT2(TX, AUX, T, IPK, KK)
C
C                PASO F23+ (LLAMA AL ALGORITMO GPALG)
C
                CALL GPALG(T, Q, K, KK, IFAULT)
                IF (IFAULT .NE. 0) RETURN
C
C                PASO F24+
C
                COS = Q(1,1)
                SIN = Q(2,1)
                DO 13 Z = 1, IP
                    B(Z, M) = H(Z, 1) * COS + H(Z, 2) * SIN
                    B(Z, N) = H(Z, 2) * COS - H(Z, 1) * SIN
13                CONTINUE
8                CONTINUE
C
C                PASO F3+
C
                EPS = ZERO
                DO 10 M = 1, IP
                    DO 10 J = 1, IP
                        DIFF = ABS(B(M,J) - BOLD(M, J))
                        IF(DIFF .GT. EPS) EPS = DIFF
10                CONTINUE
                IF(EPS .GE. EPSF .AND. NF .LT. MAXF) GOTO 4
                IF(EPS .GE. EPSF) IFAULT = 4
                RETURN
                END
C
C
C                SUBROUTINE GPALG(T, Q, K, KK, IFAULT)
                USE MSIMSL
                REAL T(KK, KK), Q(2, 2), AUX1(2,1), AUX2(2,1),
                & AUX3(KK, K), AUX4(KK, K), MTM(K, K), MTN(K, K),
                & MTMINV(K, K), MTNINV(K, K), ASTR(K,K), TSTR(2,2)
                REAL EPSG, ZERO, ONE, C, S, SOLD
                DATA ZERO /0.0/, ONE /1.0/, EPSG /1.E-4/
                MAXG = 5
                IFAULT = 0
                IF (K .EQ. 1) MAXG = 1
C
C                PASO G0+ INICIALIZAMOS Q COMO LA MATRIZ IDENTIDAD
C
                S = 0
                DO 5 I=1, 2
                    DO 5 J=1, 2
                        Q(I, J) = ZERO
                        IF (I .EQ. J) Q(I, J) = ONE
5                CONTINUE
C
C                PASO G1+ ACTUALIZAMOS S
C
                DO 1 NG = 1, MAXG
                    SOLD = S
6                CONTINUE
C

```

```

C          PASO G2+ CALCULAMOS MTM Y MTN
C
DO 7 I = 1, 2
    AUX1(I,1) = Q(I,1)
    AUX2(I,1) = Q(I,2)
7  CONTINUE
    CALL KRONI(K, AUX1, 2, 1, AUX3)
    CALL KRONI(K, AUX2, 2, 1, AUX4)
    CALL MULT2(T, AUX3, MTM, KK, K)
    CALL MULT2(T, AUX4, MTN, KK, K)
C
C          PASO G3+ GENERAMOS LA MATRIZ A* Y LA MATRIZ T*
C
CALL LINRG(K, MTM, K, MTMINV, K)
CALL LINRG(K, MTN, K, MTNINV, K)
DO 8 I = 1, K
    DO 8 J = 1, K
        ASTR(I,J)=MTMINV(I,J)-MTNINV(I,J)
8  CONTINUE
    DO 9 I = 1, 2
        DO 9 J = 1, 2
            TSTR(I,J) = ZERO
9  CONTINUE
    DO 10 I = 1, K
        DO 10 J = 1, K
            TSTR(1, 1) = ASTR(I, J) * T(2 * I - 1, 2 * J - 1) + TSTR(1, 1)
            TSTR(1, 2) = ASTR(I, J) * T(2 * I - 1, 2 * J) + TSTR(1, 2)
            TSTR(2, 1) = ASTR(I, J) * T(2 * I, 2 * J - 1) + TSTR(2, 1)
            TSTR(2, 2) = ASTR(I, J) * T(2 * I, 2 * J) + TSTR(2, 2)
10 CONTINUE
C
C          PASO G4+
C
CALL EIGVEC(TSTR, Q)
C = Q(1, 1)
S = Q(2, 1)
C
C          PASO G5+ REVISAMOS LA CONVERGENCIA
C
IF (ABS(S - SOLD) .LT. EPSG) RETURN
1 CONTINUE
RETURN
END
C
C
C          SUBROUTINE EIGVEC(U, Q)
REAL U(2, 2), Q(2, 2)
REAL ZERO, ONE, TWO, FOUR, RATIO, DISCR, C, S, T1, T2, T, EP
DATA ZERO /0.0/, ONE /1.0/, TWO /2.0/, FOUR /4.0/, EP /1.E-10/
C
C          CALCULA LOS VECTORES PROPIOS DE U EN Q
C
C = ONE
S = ZERO
IF (ABS(U(1, 2)) .LT. EP) GOTO 1
RATIO = (U(2, 2) - U(1, 1)) / U(1, 2)

```

```

DISCR = SQRT(RATIO * RATIO + FOUR)
T1 = (RATIO + DISCR) / TWO
T2 = (RATIO - DISCR) / TWO
T = T1
IF (ABS(T1) .GT. ABS(T2)) T = T2
C = ONE / SQRT(ONE + T * T)
S = T * C
1   Q(1, 1) = C
    Q(2, 1) = S
    Q(1, 2) = -S
    Q(2, 2) = C
    RETURN
    END
C
C
C
SUBROUTINE MULT2(AA, B, C, IP, IQ)
REAL AA(IP, IP), B(IP, IQ), C(IQ, IQ), BT(IQ, IP), H(IQ, IP)
REAL ZERO
DATA ZERO /0.0/
C
C      CALCULA EL PRODUCTO DE MATRICES B' AA B EN C
C
DO 4 I = 1, IQ
  DO 4 J = 1, IP
    BT(I, J) = B(J, I)
4   DO 1 I = 1, IQ
    DO 1 J = 1, IP
      H(I, J) = ZERO
      DO 1 K = 1, IP
        H(I, J) = H(I, J) + BT(I, K) * AA(K, J)
1   CONTINUE
DO 2 I = 1, IQ
  DO 2 J = 1, IQ
    C(I, J) = ZERO
    DO 3 K = 1, IP
      C(I, J) = C(I, J) + H(I, K) * B(K, J)
3   CONTINUE
2   RETURN
    END
C
C
C
SUBROUTINE KRONI(K, MT, IP, IQ, T)
REAL MT(IP, IQ), T(K*IP, K*IQ)
REAL ZERO
DATA ZERO /0.0/
C
C      CALCULA EL PRODUCTO DE KRONECKER DE LA MATRIZ Ik Y LA MATRIZ MT
C
IPK = K * IP
IQK = K * IQ
DO 1 I = 1, IPK
  DO 1 J = 1, IQK
    T(I, J) = ZERO
1   CONTINUE
DO 2 L = 1, K

```



```

        DO 2 I = 1, IP
            DO 2 J = 1, IQ
                T(IP*(L-1)+I, IQ*(L-1)+J) = MT(I, J)
2      CONTINUE
      RETURN
      END
C
C
C
      SUBROUTINE KRON(A, AP, AQ, B, BP, BQ, T)
      INTEGER AP, AQ, BP, BQ
      REAL A(AP, AQ), B(BP, BQ), T(AP*BP, AQ*BQ)
C
C      CALCULA EL PRODUCTO DE KRONECKER DE LA MATRIZ A
C      DE DIMENSION AP X AQ Y LA MATRIZ B DE DIMENSION BP X BQ
C
      DO 1 I = 1, AP
          DO 1 J = 1, AQ
              DO 1 L = 1, BP
                  DO 1 M=1,BQ
                      T(L+(BP*(I-1)),M+(BQ*(J-1)))= A(I,J)*B(L,M)
1      CONTINUE
      END
C
C
C
      SUBROUTINE DETERM(F, IPK, DETF)
      REAL F(IPK, IPK), FACF(IPK, IPK)
      INTEGER IPVTF(IPK)
      REAL DETF, DETF1, DETF2
C
C      CALCULA EL DETERMINANTE DE LA MATRIZ F
C      DE DIMENSION IPK X IPK
C
      CALL LFTRG (IPK, F, IPK, FACF, IPK, IPVTF)
      CALL LFDRG (IPK, FACF, IPK, IPVTF, DETF1, DETF2)
      DETF=DETF1*(10**DETF2)
      END
C
C
C
      SUBROUTINE BLOQDIAG(F, IP, K, IPK, LAMBDA)
      REAL F(IPK, IPK), LAMBDA(IPK, IPK)
      INTEGER IP, K
C
C      CALCULA LA MATRIZ DIAGONAL POR BLOQUES LAMBDA DE LA MATRIZ F
C      DE DIMENSION IPK X IPK
      DO 1 I = 1, K
          DO 1 J = 1, K
              DO 1 L = 1, IP
                  DO 1 M=1,IP
                      IF (L .EQ. M) THEN
                          LAMBDA(L+(IP*(I-1)),M+(IP*(J-1)))=
&                          F(L+(IP*(I-1)),M+(IP*(J-1)))
                      ELSE
                          LAMBDA(L+(IP*(I-1)),M+(IP*(J-1)))= 0
                      ENDIF

```

```

1  CONTINUE
   END
C
C
C
SUBROUTINE JACOB(B, P, J1, J2, J3)
INTEGER P, REN, COLA, COLC, Q
REAL B(P, P), M(P*(P+1)/2,P*P)
&      , J1(P*(P-1)/2,P*P), J2(P*(P-1)/2,P*P),
&      J3(P*(P-1)/2,P*P)
&      , A(P*(P+1)/2, P*(P+1)/2), C(P*(P+1)/2, P*(P-1)/2),
&      IND(P*(P-1)/2, P*(P-1)/2),
&      D(P*(P+1)/2, P*(P-1)/2), AINV(P*(P+1)/2, P*(P+1)/2)
C
C      CALCULA LOS TRES JACOBIANOS NECESARIOS PARA ESTIMAR
C      LA MATRIZ DE VARIANZAS Y COVARIANZAS DE BETA DE DIMENSION
C      TRES O MAS
C
REN=0
DO 1 I = 1, P
  DO 1 J = I, P
    REN=REN+1
    IF (I .EQ. J) THEN
      DO 2 K=1,P*(I-1)
2         M(REN,K)= 0
      DO 3 K= P*(I-1)+1, P*(I)
3         M(REN,K)= 2*(B(K- P*(I-1),I))
      DO 4 K= P*(I)+1, P*P
4         M(REN,K)= 0
      ENDIF
      IF (I .NE. J) THEN
        DO 5 K=1,P*(I-1)
5         M(REN,K)= 0
        DO 6 K= P*(I-1)+1, P*(I)
6         M(REN,K)= (B(K- P*(I-1),J))
        DO 7 K= P*(I)+1, P*(J-1)
7         M(REN,K)= 0
        DO 8 K= P*(J-1)+1, P*(J)
8         M(REN,K)= (B(K- P*(J-1),I))
        DO 9 K= P*(J)+1, P*P
9         M(REN,K)= 0
      ENDIF
1  CONTINUE
C
COLA=1
COLC=1
REN=0
DO 10 I = 1, P
  REN=REN+1
  DO 10 J = 1, P
    IF (J .LE. REN) THEN
      DO 11 K = 1, P*(P+1)/2
11         A(K,COLA)=M(K,J+ P*(REN-1))
      CONTINUE
      COLA=COLA+1
    ENDIF

```

```

                IF (J .GT. REN) THEN
                    DO 12 K = 1, P*(P+1)/2
                        C(K, COLC)=M(K, J+ P*(REN-1))
12                CONTINUE
                    COLC=COLC+1
                ENDIF
10            CONTINUE
C
        CALL LINRG(P*(P+1)/2, A, P*(P+1)/2, AINV, P*(P+1)/2)
        D=-MATMUL(AINV,C)
C
        DO 13 I=1,P*(P-1)/2
            DO 13 J=1,P*(P-1)/2
                IF (I.EQ.J) THEN
                    IND(I,J)=1
                ELSE
                    IND(I,J)=0
                ENDIF
13            CONTINUE
C
        COLA=1
        COLC=1
        REN=0
        DO 14 I = 1, P
            REN=REN+1
            DO 14 J = 1, P
                IF (J .LE. REN) THEN
                    DO 15 K = 1, P*(P-1)/2
                        J1(K, J+ P*(REN-1))=D(COLA, K)
15                CONTINUE
                    COLA=COLA+1
                ENDIF
                IF (J .GT. REN) THEN
                    DO 16 K = 1, P*(P-1)/2
                        J1(K, J+ P*(REN-1))=IND(COLC, K)
16                CONTINUE
                    COLC=COLC+1
                ENDIF
14            CONTINUE
C
        COLA=1
        COLC=1
        REN=0
        DO 20 I = 1, P
            REN=REN+1
            DO 20 J = 1, P
                IF (J .GE. REN) THEN
                    DO 21 K = 1, P*(P+1)/2
                        A(K, COLA)=M(K,          J+ P*(REN-1))
21                CONTINUE
                    COLA=COLA+1
                ENDIF
                IF (J .LT. REN) THEN
                    DO 22 K = 1, P*(P+1)/2
                        C(K, COLC)=M(K,          J+ P*(REN-1))
22                CONTINUE
                    COLC=COLC+1
                ENDIF

```

```

20   CONTINUE
C
CALL LINRG(P*(P+1)/2, A, P*(P+1)/2, AINV, P*(P+1)/2)
D=-MATMUL(AINV,C)
C
DO 23 I=1,P*(P-1)/2
    DO 23 J=1,P*(P-1)/2
        IF (I.EQ.J) THEN
            IND(I,J)=1
        ELSE
            IND(I,J)=0
        ENDIF
23   CONTINUE
C
COLA=1
COLC=1
REN=0
DO 24 I = 1, P
    REN=REN+1
    DO 24 J = 1, P
        IF (J .GE. REN) THEN
            DO 25 K = 1, P*(P-1)/2
                J2(K,J+ P*(REN-1))=D(COLA,K)
25                CONTINUE
                COLA=COLA+1
            ENDIF
            IF (J .LT. REN) THEN
                DO 26 K = 1, P*(P-1)/2
                    J2(K,J+ P*(REN-1))=IND(COLC,K)
26                CONTINUE
                    COLC=COLC+1
                ENDIF
24   CONTINUE
C
COLC=1
DO 30 I = 1, P
    DO 30 J = 1, P
        IF (J .EQ. I) THEN
            DO 32 K = 1, P*(P+1)/2
                C(K,COLC)=M(K,J+ P*(I-1))
32   CONTINUE
                COLC=COLC+1
            ENDIF
30   CONTINUE
C
Q= INT(P*(P-1)/2)
COLA=1
REN=0
DO 40 I = 1, P-1
    DO 40 J = I+1, P
        IF (COLC .GT. Q) THEN
            DO 42 K = 1, P*(P+1)/2
                A(K,COLA)=M(K,J+ P*(J-I-1))
42                CONTINUE
                    COLA=COLA+1
                ENDIF
C
                IF (COLC .LE. Q) THEN

```

```

DO 41 K = 1, P*(P+1)/2
    C(K, COLC)=M(K, J+ P*(J-I-1))
41    CONTINUE
    COLC=COLC+1
    ENDIF
C
    IF (COLC .GT. Q) THEN
        DO 52 K = 1, P*(P+1)/2
            A(K, COLA)=M(K, (J-I)+ P*(J-1))
52    CONTINUE
        COLA=COLA+1
        ENDIF
C
        IF (COLC .LE. Q) THEN
            DO 51 K = 1, P*(P+1)/2
                C(K, COLC)=M(K, (J-I)+ P*(J-1))
51    CONTINUE
            COLC=COLC+1
            ENDIF
C
40    CONTINUE
C
CALL LINRG(P*(P+1)/2, A, P*(P+1)/2, AINV, P*(P+1)/2)
C
D=-MATMUL(AINV,C)
C
DO 33 I=1,P*(P-1)/2
    DO 33 J=1,P*(P-1)/2
        IF (I.EQ.J) THEN
            IND(I,J)=1
        ELSE
            IND(I,J)=0
        ENDIF
33    CONTINUE
C
COLC=1
REN=0
DO 44 I = 1, P
    REN=REN+1
    DO 44 J = 1, P
        IF (J .EQ. REN) THEN
            DO 46 K = 1, P*(P-1)/2
                J3(K, J+ P*(REN-1))=IND(COLC, K)
46    CONTINUE
            COLC=COLC+1
        ENDIF
44    CONTINUE
C
COLA=1
DO 47 I = 1, P-1
    DO 47 J = I+1, P
C
        IF (COLC .GT. Q) THEN
            DO 48 K = 1, P*(P-1)/2
                J3(K, J+ P*(J-I-1))=D(COLA, K)
48    CONTINUE
            COLA=COLA+1
        ENDIF

```

```

C          IF (COLC .LE. Q) THEN
              DO 49 K = 1, P*(P-1)/2
                  J3(K,J+ P*(J-1))=IND(COLC,K)
49          CONTINUE
              COLC=COLC+1
          ENDIF
C          IF (COLC .GT. Q) THEN
              DO 53 K = 1, P*(P-1)/2
                  J3(K,(J-1)+ P*(J-1))=D(COLA,K)
53          CONTINUE
              COLA=COLA+1
          ENDIF
C          IF (COLC .LE. Q) THEN
              DO 54 K = 1, P*(P-1)/2
                  J3(K,(J-1)+ P*(J-1))=IND(COLC,K)
54          CONTINUE
              COLC=COLC+1
          ENDIF
C
47      CONTINUE
      END
C
C
C
SUBROUTINE VARBNORM(J1, J2, J3, LAMBDA, B, IP, K ,IPK, ENE, VB)
INTEGER IP,K, IPK, REN, ENE
REAL LAMBDA(IPK,IPK), LAMBDAINV(IPK,IPK), B(IP,IP),
&      J1(IP*(IP-1)/2,IP*IP),J2(IP*(IP-1)/2,IP*IP),
&      J3(IP*(IP-1)/2,IP*IP),
&      LIJ(IP,IP), LIJINV(IP,IP), S(IP*IP,IP*IP), LKRONL(IP*IP,IP*IP),
&      KIP2(IP*IP,IP*IP), IPKRONB(IP*IP,IP*IP), SS(IP*IP,IP*IP),
&      INF(IP*(IP-1)/2,IP*(IP-1)/2), INFINV(IP*(IP-1)/2,IP*(IP-1)/2),
&      VB(IP,IP)
C
C      CALCULA LA LA MATRIZ DE VARIANZAS Y COVARIANZAS ESTIMADA DE BETA
C      PARA EL SUPUESTO DE NORMALIDAD, PARA DIMENSION
C      TRES O MAS
C
      DO 15 I=1,IP*IP
          DO 15 J=1,IP*IP
              S(I,J)=0
15      CONTINUE
C
      CALL LINRG(IPK, LAMBDA, IPK, LAMBDAINV, IPK)
      DO 16 I=1,K
          DO 16 J=1,K
              DO 17 L=1,IP
                  DO 17 M=1,IP
                      LIJ(L,M)=LAMBDA(L+(I-1)*IP,M+(J-1)*IP)
                      LIJINV(L,M)=LAMBDAINV(L+(I-1)*IP,M+(J-1)*IP)
17          CONTINUE
              CALL KRON(LIJINV, IP, IP, LIJ, IP, IP, LKRONL)
              S=LKRONL+S
16      CONTINUE

```

```

C
DO 18 I=1,IP*IP
  DO 18 J=1,IP*IP
    IF (I .EQ. J) THEN
      KIP2(I,J)=K
    ELSE
      KIP2(I,J)=0
    ENDIF
18  CONTINUE
C
S=S-KIP2
C
CALL KRONI(IP, B, IP, IP, IPKRONB)
CALL MULT2(S, TRANSPOSE(IPKRONB), SS, IP*IP, IP*IP)
CALL MULT2(SS, TRANSPOSE(J1), INF, IP*IP, IP*(IP-1)/2)
INF=ENE*INF
CALL LINRG(IP*(IP-1)/2, INF, IP*(IP-1)/2, INFINV, IP*(IP-1)/2)
C
REN=1
DO 19 I=1,IP
  DO 19 J=I+1,IP
    VB(J,I)=INFINV(REN,REN)
    REN=REN+1
19  CONTINUE
C
CALL MULT2(SS, TRANSPOSE(J2), INF, IP*IP, IP*(IP-1)/2)
INF=ENE*INF
CALL LINRG(IP*(IP-1)/2, INF, IP*(IP-1)/2, INFINV, IP*(IP-1)/2)
C
REN=1
DO 20 I=2,IP
  DO 20 J=1,I-1
    VB(J,I)=INFINV(REN,REN)
    REN=REN+1
20  CONTINUE
C
CALL MULT2(SS, TRANSPOSE(J3), INF, IP*IP, IP*(IP-1)/2)
INF=ENE*INF
CALL LINRG(IP*(IP-1)/2, INF, IP*(IP-1)/2, INFINV, IP*(IP-1)/2)
C
REN=1
DO 21 I=1,IP
  VB(I,I)=INFINV(REN,REN)
  REN=REN+1
21  CONTINUE
END
C
C
C
SUBROUTINE JACOB2(B, P, J1, J2)
INTEGER P, REN, COLA
REAL B(P, P), M(P*(P+1)/2,P*P)
&    , J1(P*(P-1)/2,P*P), J2(P*(P-1)/2,P*P)
&    , A(P*(P+1)/2, P*(P+1)/2), C(P*(P+1)/2, P*(P-1)/2),
&    D(P*(P+1)/2, P*(P-1)/2), AINV(P*(P+1)/2, P*(P+1)/2)
C
C    CALCULA LOS DOS JACOBIANOS NECESARIOS PARA ESTIMAR

```

```

C          LA MATRIZ DE VARIANZAS Y COVARIANZAS DE BETA DE DIMENSION
C          DOS
C
REN=0
DO 1 I = 1, P
  DO 1 J = I, P
    REN=REN+1
    IF (I .EQ. J) THEN
      DO 2 K=1,P*(I-1)
        M(REN,K)= 0
      DO 3 K= P*(I-1)+1, P*(I)
        M(REN,K)= 2*(B(K- P*(I-1),I))
      DO 4 K= P*(I)+1, P*P
        M(REN,K)= 0
    ENDIF
    IF (I .NE. J) THEN
      DO 5 K=1,P*(I-1)
        M(REN,K)= 0
      DO 6 K= P*(I-1)+1, P*(I)
        M(REN,K)= (B(K- P*(I-1),J))
      DO 7 K= P*(I)+1, P*(J-1)
        M(REN,K)= 0
      DO 8 K= P*(J-1)+1, P*(J)
        M(REN,K)= (B(K- P*(J-1),I))
      DO 9 K= P*(J)+1, P*P
        M(REN,K)= 0
    ENDIF
1 CONTINUE
C
COLA=1
REN=0
DO 10 I = 1, P
  REN=REN+1
  DO 10 J = 1, P
    IF (I .EQ. 1 .AND. J .EQ. 1) THEN
      DO 11 K = 1, P*(P+1)/2
        C(K,1)=M(K,J+ P*(REN-1))
11      CONTINUE
      COLC=COLC+1
    ELSE
      DO 12 K = 1, P*(P+1)/2
        A(K,COLA)=M(K,J+ P*(REN-1))
12      CONTINUE
      COLA=COLA+1
    ENDIF
10 CONTINUE
C
CALL LINRG(P*(P+1)/2, A, P*(P+1)/2, AINV, P*(P+1)/2)
D=-MATMUL(AINV,C) C
13 CONTINUE
C
COLA=1
REN=0
DO 14 I = 1, P
  REN=REN+1
  DO 14 J = 1, P

```



```

                IF (I .EQ. 1 .AND. J .EQ. 1) THEN
                    DO 15 K = 1, P*(P-1)/2
                        J1(K,J+ P*(REN-1))=1
15                CONTINUE
                ELSE
                    DO 16 K = 1, P*(P-1)/2
                        J1(K,J+ P*(REN-1))=D(COLA,K)
16                CONTINUE
                    COLA=COLA+1
                ENDIF
14    CONTINUE
C
COLA=1
REN=0
DO 20 I = 1, P
    REN=REN+1
    DO 20 J = 1, P
        IF (I .EQ. 1 .AND. J .EQ. 2) THEN
            DO 21 K = 1, P*(P+1)/2
                C(K,1)=M(K,J+ P*(REN-1))
21            CONTINUE
            ELSE
                DO 22 K = 1, P*(P+1)/2
                    A(K,COLA)=M(K,J+ P*(REN-1))
22                CONTINUE
                    COLA=COLA+1
                ENDIF
20    CONTINUE
C
CALL LINRG(P*(P+1)/2, A, P*(P+1)/2, AINV, P*(P+1)/2)
D=-MATMUL(AINV,C)
C
COLA=1
REN=0
DO 24 I = 1, P
    REN=REN+1
    DO 24 J = 1, P
        IF (I .EQ. 1 .AND. J .EQ. 2) THEN
            DO 25 K = 1, P*(P-1)/2
                J2(K,J+ P*(REN-1))=1
25            CONTINUE
            ELSE
                DO 26 K = 1, P*(P-1)/2
                    J2(K,J+ P*(REN-1))=D(COLA,K)
26                CONTINUE
                    COLA=COLA+1
                ENDIF
24    CONTINUE
END
C
C
C
C
SUBROUTINE VARBNORM2(J1, J2, LAMBDA, B, IP, K ,IPK, ENE, VB)
INTEGER IP,K, IPK, ENE
REAL LAMBDA(IPK,IPK), LAMBDAINV(IPK,IPK), B(IP,IP),
&    J1(IP*(IP-1)/2,IP*IP), J2(IP*(IP-1)/2,IP*IP),
&    LIJ(IP,IP), LIJINV(IP,IP), S(IP*IP,IP*IP), LKRONL(IP*IP,IP*IP),

```

```

&      KIP2(IP*IP,IP*IP), IPKRONB(IP*IP,IP*IP), SS(IP*IP,IP*IP),
&      INF(IP*(IP-1)/2,IP*(IP-1)/2), INFINV(IP*(IP-1)/2,IP*(IP-1)/2),
&      VB(IP,IP)
C
C      CALCULA LA LA MATRIZ DE VARIANZAS Y COVARIANZAS ESTIMADA DE BETA
C      PARA EL SUPUESTO DE NORMALIDAD, PARA DIMENSION
C      TRES O MAS
C
DO 15 I=1,IP*IP
      DO 15 J=1,IP*IP
          S(I,J)=0
15  CONTINUE
C
CALL LINRG(IPK, LAMBDA, IPK, LAMBDAINV, IPK)
DO 16 I=1,K
      DO 16 J=1,K
          DO 17 L=1,IP
              DO 17 M=1,IP
                  LIJ(L,M)=LAMBDA(L+(I-1)*IP,M+(J-1)*IP)
                  LIJINV(L,M)=LAMBDAINV(L+(I-1)*IP,M+(J-1)*IP)
17          CONTINUE
          CALL KRON(LIJINV, IP, IP, LIJ, IP, IP, LKRONL)
          S=LKRONL+S
16  CONTINUE
C
DO 18 I=1,IP*IP
      DO 18 J=1,IP*IP
          IF (I .EQ. J) THEN
              KIP2(I,J)=K
          ELSE
              KIP2(I,J)=0
          ENDIF
18  CONTINUE
C
S=S-KIP2
CALL KRONI(IP, B, IP, IP, IPKRONB)
CALL MULT2(S, TRANSPOSE(IPKRONB), SS, IP*IP, IP*IP)
CALL MULT2(SS, TRANSPOSE(J1), INF, IP*IP, IP*(IP-1)/2)
INF=ENE*INF
CALL LINRG(IP*(IP-1)/2, INF, IP*(IP-1)/2, INFINV, IP*(IP-1)/2)
C
DO 19 I=1,IP
      DO 19 J=1,IP
          IF (I .EQ. J) THEN
              VB(J,I)=INFINV(1,1)
          ENDIF
19  CONTINUE
C
CALL MULT2(SS, TRANSPOSE(J2), INF, IP*IP, IP*(IP-1)/2)
INF=ENE*INF
CALL LINRG(IP*(IP-1)/2, INF, IP*(IP-1)/2, INFINV, IP*(IP-1)/2)
C
DO 20 I=1,IP
      DO 20 J=1,IP
          IF (I .NE. J) THEN
              VB(J,I)=INFINV(1,1)

```

```

                ENDIF
20    CONTINUE
C
END
C
C
C
SUBROUTINE MATVARCOVA(P, N, SIMUL, VARCOV)
USE MSIMSL
C
C    PROGRAMA EN FORTRAN PARA CALCULAR LA MATRIZ DE
C    VARIANZAS Y COVARIANZAS MUESTRAL (W) DE UN CONJUNTO
C    DE DATOS
C
C    INTEGER P, N
C    REAL SIMUL(N, P), EIND(1,P), ECONJ(P,P), VARCOV(P, P)
C
C    DO 1 I = 1, P
C        EIND(1,I) = 0
C        DO 2 J = 1, P
C            ECONJ (J,I) = 0
C            VARCOV(P,P) = 0
2    CONTINUE
1    CONTINUE
C
C    DO 5 I = 1, N
C        DO 3 J = 1, P
C            EIND(1,J) = SIMUL(I,J) + EIND(1,J)
C            DO 4 K = 1, P
C                ECONJ (K,J) = SIMUL(I,K) * SIMUL(I,J) + ECONJ(K,J)
4    CONTINUE
3    CONTINUE
5    CONTINUE
C
C    DO 6 I = 1, P
C        DO 6 J = 1, P
C            VARCOV(I,J) = (ECONJ(I,J)/N)-((EIND(1,I)*EIND(1,J))/(N*N))
6    CONTINUE
END
C
C
C
SUBROUTINE SIMNORM(MU, SIGMA, P, N, SIMUL)
USE MSIMSL
USE PORTLIB
C
C    PROGRAMA EN FORTRAN PARA SIMULAR N VECTORES P-VARIADOS
C    DE UNA NORMAL CON:
C    MU=VECTOR DE MEDIAS
C    SIGMA=MATRIZ DE VARIANZAS COVARIANZAS P X P (INPUT)
C    SIMUL=MATRIZ DE DATOS SIMULADOS N X P (OUTPUT)
C
C    INTEGER P, N
C    REAL A, B, C, ACCEPT, U, V, W, X, Y, Z, R
C    REAL MU(1,P), SIGMA(P, P), SIMUL(N, P), AA(P, P)
C    REAL THETA, UN(1,P), AUXILIAR(1,P), SR(P,P), VR(P,P)
C    COMPLEX SS(P), VS(P,P)
C

```

```

C          OBTIENE LA DESCOMPOSICION SVD DE SIGMA
C
CALL EVCRG (P, SIGMA, P, SS, VS, P)
C
DO 15 I = 1 , P
    DO 15 J = 1 , P
        VR(I,J)=REAL(VS(I,J))
15         SR(I,J)=0 C
DO 16 I = 1 , P
16         SR(I,I)=REAL(SS(I)**0.5
C
AA=MATMUL(VR,SR)
C
C          COMIENZA CON EL CICLO DE SIMULACIONES
C
DO 1 M = 1, N
C
C          GENERA UN VECTOR UNIFORME EN LA ESFERA UNITARIA
C          UTILIZANDO COORDENADAS ESFERICAS
C
ACCEPT = 0
DO WHILE (ACCEPT .EQ. 0)
    THETA = 0
    DO 2 I = 1, P
        UN(1,I) = RAND()*2 -1
        THETA = THETA + UN(1,I)*UN(1,I)
2    CONTINUE
    THETA= THETA ** 0.5
    IF (THETA .LE. 1) THEN
        ACCEPT=1
    ENDIF
END DO
DO 3 I = 1, P
3    UN(1,I) = UN(1,I)/THETA
C
C          GENERA UN VALOR DE R
C
A = P / 2
B = A - 1
C = 3 * A - (3 / 4)
ACCEPT = 0
DO WHILE (ACCEPT .EQ. 0)
    U=RAND()
    V=RAND()
    W=U * (1-U)
    Y = ((C / W) ** 0.5) * (U - 0.5)
    X = B + Y
    IF (X .GE. 0) THEN
        Z = 64 * (W ** 3) * (V ** 2)
        IF (Z .LE. (1-((2 * (Y ** 2)) / X))) THEN
            ACCEPT = 1
        ELSE IF ((B .GT. 0) .AND.
&            (LOG(Z) .LE. (2 * (B * LOG(X / B) - Y)))) THEN
            ACCEPT = 1
        ELSE IF ((B .EQ. 0) .AND. (LOG(Z) .LE. (-2 * Y))) THEN
            ACCEPT=1
        END IF
    END IF
END

```

```

END      DO
R=RAND()*2-1
R=R/(ABS(R))
R = R*(2 * X) ** 0.5

C
C      OBTIENE EL VECTOR SIMULADO
C
AUXILIAR = TRANSPOSE(TRANSPOSE(MU)
&          + R * MATMUL(AA,TRANSPOSE(UN)))
DO 4 L = 1, P
4      SIMUL(M,L) = AUXILIAR(1,L)
1      CONTINUE
      RETURN
      END

C
C
C
SUBROUTINE SIMPEAR(MU, SIGMA, GAM, EME, P, N, SIMUL)
USE MSIMSL
USE PORTLIB

C
C      PROGRAMA EN FORTRAN PARA SIMULAR N VECTORES P-VARIADOS
C      DE UNA PEARSON TIPO VII CON:
C      MU=VECTOR DE MEDIAS
C      SIGMA=MATRIZ DE VARIANZAS COVARIANZAS P X P (INPUT)
C      PARAMETROS GAM Y EME (INPUT)
C      SIMUL=MATRIZ DE DATOS SIMULADOS N X P (OUTPUT)
C
INTEGER P, N
REAL A1, A2, B1, B2, C1, C2, ACCEPT
REAL U, V, W, X1, X2, Y, Z, R, GAM, EME
REAL MU(1,P), SIGMA(P, P), SIMUL(N, P), AA(P, P)
REAL THETA, UN(1,P), AUXILIAR(1,P), SR(P,P), VR(P,P)
COMPLEX SS(P), VS(P,P)

C
C      OBTIENE LA DESCOMPOSICION SVD DE SIGMA
C
CALL EVCRG (P, SIGMA, P, SS, VS, P)

C
DO 15 I = 1 , P
      DO 15 J = 1 , P
          VR(I,J)=REAL(VS(I,J))
15          SR(I,J)=0
C
DO 16 I = 1 , P
16          SR(I,I)=REAL(SS(I))**0.5
C
AA=MATMUL(VR,SR)

C
C      COMIENZA CON EL CICLO DE SIMULACIONES
C
DO 1 M = 1, N

C
C      GENERA UN VECTOR UNIFORME EN LA ESFERA UNITARIA
C      UTILIZANDO COORDENADAS ESFERICAS
C
ACCEPT = 0

```

```

DO WHILE (ACCEPT .EQ. 0)
THETA = 0
  DO 2 I = 1, P
    UN(1,I) = RAND()*2 -1
    THETA = THETA + UN(1,I)*UN(1,I)
2    CONTINUE
    THETA= THETA ** 0.5
    IF (THETA .LE. 1) THEN
      ACCEPT=1
    ENDIF
  END DO
DO 3 I = 1, P
3    UN(1,I) = UN(1,I)/THETA
C
C    GENERA UN VALOR DE R
C
A1 = P / 2
B1 = A1 - 1
C1= 3 * A1 - (3 / 4)
C
A2 = (2 * EME - P) / 2
B2 = A2 - 1
C2 = 3 * A2 - (3 / 4)
C
ACCEPT = 0
DO WHILE (ACCEPT .EQ. 0)
  U = RAND()
  V = RAND()
  W = U * (1 - U)
  Y = ((C1 / W) ** 0.5) * (U - 0.5)
  X1 = B1 + Y
  IF (X1 .GE. 0) THEN
    Z = 64 * (W ** 3) * (V ** 2)
    IF (Z .LE. (1 - ((2 * (Y ** 2)) / X1))) THEN
      ACCEPT = 1
    ELSEIF ((B1 .GT. 0) .AND.
&          (LOG(Z) .LE. (2* (B1 * LOG(X1 / B1) - Y)))) THEN
      ACCEPT = 1
    ELSEIF ((B1 .EQ. 0) .AND. (LOG(Z) .LE. (-2 * Y))) THEN
      ACCEPT = 1
    END IF
  END IF
END DO
C
ACCEPT = 0
DO WHILE (ACCEPT .EQ. 0)
  U = RAND()
  V = RAND()
  W = U * (1 - U)
  Y = ((C2 / W) ** 0.5) * (U - 0.5)
  X2 = B2 + Y
  IF (X2 .GE. 0) THEN
    Z = 64 * (W ** 3) * (V ** 2)
    IF (Z .LE. (1 - ((2 * (Y ** 2)) / X2))) THEN
      ACCEPT = 1
    ELSEIF ((B2 .GT. 0) .AND.
&          (LOG(Z) .LE. (2* (B2 * LOG(X2 / B2) - Y)))) THEN
      ACCEPT = 1

```

```

                ELSEIF ((B2 .EQ. 0) .AND. (LOG(Z) .LE. (-2 * Y))) THEN
                    ACCEPT = 1
                END IF
            END IF
        END DO
        R=RAND()*2-1
        R=R/(ABS(R))
        R=R * (((GAM * P) / (2 * EME - p)) *
&          (2 * X1 / P) / (2 * X2 / (2 * EME - P))) ** 0.5
C
C          OBTIENE EL VECTOR SIMULADO
C
        AUXILIAR = TRANSPOSE(TRANSPOSE(MU)
&          + R * MATMUL(AA,TRANSPOSE(UN)))
        DO 4 L = 1, P
4          SIMUL(M,L) = AUXILIAR(1,L)
1          CONTINUE
        RETURN
        END
C
C
C
        SUBROUTINE GRUPALEAT(SIMUL, GRUP, IPK, IP, K, N, VAREST)
        USE MSIMSL
        USE PORTLIB
C
C          PROGRAMA EN FORTRAN PARA CALCULAR LA
C          MATRIZ DE VARIANZAS MUESTRALES DE LA
C          ENTRADAS DE LA MATRIZ BETA RESULTANTE
C          DE UN ANALISIS DE COMPONENTES PRINCIPALES
C          COMUNES EN VECTORES DEPENDIENTES
C
        INTEGER IPK, IP, K, N, GRUP, IFAULT, NF, L,
&          ACCEPT, AUX1
        INTEGER NPORGRUP(1,GRUP),ASIGNA(1,N)
        REAL VAREST(IP, IP), SIMUL(N, IPK), VARCOV(IPK,IPK),
&          B(IP,IP), B2(IP,IP), CAMBIA, CHECA, CHECA2
        REAL, POINTER :: AUXI(:,:)
        DATA ZERO /0.0/
C
        DO 1 I = 1, GRUP
1          NPORGRUP(1,I)= 0
C
        AUX1=0
        DO 2 I = 1, N
            IF (AUX1 .EQ. GRUP) THEN
                AUX1=0
            ENDIF
            AUX1 = 1+AUX1
            ASIGNA(1,I) =AUX1
            NPORGRUP(1,AUX1)= NPORGRUP(1,AUX1)+1
2          CONTINUE
C
        CALL MATVARCOVA(IPK, N, SIMUL, VARCOV)
C
        CALL FGPA LG(VARCOV, B, IP, K, IPK, IFAULT, NF)
C
        DO 3 I= 1, IP

```

```

DO 3 J=1, IP
3      VAREST(I,J)=0
C
DO 4 I = 1, GRUP
    AUX1=NPORGRUP(1,I)
    ALLOCATE (AUXI(AUX1, IPK))
    L=1
        DO 5 J = 1, AUX1
            ACCEPT = 0
            DO WHILE ((ACCEPT .EQ. 0) .AND. (L .LE. N))
                IF (ASIGNA(1,L) .EQ. I) THEN
                    AUXI(J,:)=SIMUL(L,:)
                    L=L+1
                    ACCEPT =1
                ELSE
                    L=L+1
                END IF
            END DO
5        CONTINUE
        CALL MATVARCOVA(IPK, AUX1, AUXI, VARCOV)
        CALL FGPALG(VARCOV, B2, IP, K, IPK, IFAULT, NF)
C
        DO 8 L= 1, IP
            CHECA2=0
            DO 8 J=L, IP
                CHECA=0
                DO 9 H=1,IP
                    CHECA=CHECA+ABS(ABS(B2(H,J))-ABS(B(H,L)))
9                CONTINUE
                IF (CHECA2 .EQ. ZERO) THEN
                    CHECA2=CHECA
                ENDIF
                IF (CHECA .LT. CHECA2) THEN
                    CHECA2=CHECA
                    DO 10 H=1,IP
                        CAMBIA=B2(H,L)
                        B2(H,L)=B2(H,J)
                        B2(H,J)=CAMBIA
                        IF (B2(H,L)*B(H,L) .LE. ZERO) THEN
                            B2(H,L)=-B2(H,L)
10                        ENDIF
                        CONTINUE
                    ENDIF
8                CONTINUE
            ENDIF
        CONTINUE
        DO 6 L= 1, IP
            DO 6 J=1, IP
                VAREST(L,J)=VAREST(L,J)+(B2(L,J)-B(L,J))**2
6            CONTINUE
4        CONTINUE
C
        DO 7 I= 1, IP
            DO 7 J=1, IP
7                VAREST(I,J)=VAREST(I,J)/(GRUP*(GRUP-1))
            END
C

```



```

C
C
SUBROUTINE JACKKN(SIMUL, GRUP, IPK, IP, K, N, VAREST)
USE MSIMSL
USE PORTLIB

C
C      PROGRAMA EN FORTRAN PARA CALCULAR LA
C      MATRIZ DE VARIANZAS MUESTRALES DE LA
C      ENTRADAS DE LA MATRIZ BETA RESULTANTE
C      DE UN ANALISIS DE COMPONENTES PRINCIPALES
C      COMUNES EN VECTORES DEPENDIENTES
C
INTEGER IPK, IP, K, N, GRUP, IFAULT, NF, L,
&      ACCEPT, AUX1
INTEGER NPORGRUP(1,GRUP),ASIGNA(1,N)
REAL VAREST(IP, IP), SIMUL(N, IPK), VARCOV(IPK,IPK),
&      B(IP,IP), B2(IP,IP), CAMBIA, CHECA, CHECA2
REAL, POINTER :: AUXI(:,:)
DATA ZERO /0.0/

C
DO 1 I = 1, GRUP
1   NPORGRUP(1,I)= 0
C
AUX1=0
DO 2 I = 1, N
    IF (AUX1 .EQ. GRUP) THEN
        AUX1=0
    ENDIF
    AUX1 = 1+AUX1
    ASIGNA(1,I) =AUX1
    NPORGRUP(1,AUX1)= NPORGRUP(1,AUX1)+1
2   CONTINUE
C
CALL MATVARCOVA(IPK, N, SIMUL, VARCOV)
CALL FGPALG(VARCOV, B, IP, K, IPK, IFAULT, NF)
C
DO 3 I= 1, IP
    DO 3 J=1, IP
3       VAREST(I,J)=0
C
DO 4 I = 1, GRUP
    AUX1=N-NPORGRUP(1,I)
    ALLOCATE (AUXI(AUX1, IPK))
    L=1
    DO 5 J = 1, AUX1
        ACCEPT = 0
        DO WHILE ((ACCEPT .EQ. 0) .AND. (L .LE. N))
            IF (ASIGNA(1,L) .NE. I) THEN
                AUXI(J,:)=SIMUL(L,:)
                L=L+1
                ACCEPT =1
            ELSE
                L=L+1
            END IF
        END DO
5       CONTINUE
    CALL MATVARCOVA(IPK, AUX1, AUXI, VARCOV)
    CALL FGPALG(VARCOV, B2, IP, K, IPK, IFAULT, NF)

```

```

C          DO 8 L= 1, IP
            CHECA2=0
            DO 8 J=L, IP
                CHECA=0
                DO 9 H=1,IP
                    CHECA=CHECA+ABS(ABS(B2(H,J))-ABS(B(H,L)))
9          CONTINUE
C
            IF (CHECA2 .EQ. ZERO) THEN
                CHECA2=CHECA
            ENDIF
            IF (CHECA .LT. CHECA2) THEN
                CHECA2=CHECA
                DO 10 H=1,IP
                    CAMBIA=B2(H,L)
                    B2(H,L)=B2(H,J)
                    B2(H,J)=CAMBIA
                    IF (B2(H,L)*B(H,L) .LE. ZERO) THEN
                        B2(H,L)=-B2(H,L)
10         ENDIF
                CONTINUE
            ENDIF
8          CONTINUE
            DO 6 L= 1, IP
                DO 6 J=1, IP
                    VAREST(L,J)=VAREST(L,J)+(B2(L,J)-B(L,J))**2
6          CONTINUE
4          CONTINUE
C
            DO 7 I= 1, IP
                DO 7 J=1, IP
7          VAREST(I,J)=VAREST(I,J)*(GRUP-1)/GRUP
            END
C
C
C          SUBROUTINE REPLICA(SIMUL, GRUP, IPK, IP, K, N, VAREST)
            USE MSIMSL
            USE PORTLIB
C
C          PROGRAMA EN FORTRAN PARA CALCULAR LA
C          MATRIZ DE VARIANZAS MUESTRALES DE LA
C          ENTRADAS DE LA MATRIZ BETA RESULTANTE
C          DE UN ANALISIS DE COMPONENTES PRINCIPALES
C          COMUNES EN VECTORES DEPENDIENTES
C
            INTEGER IPK, IP, K, N, GRUP, IFAULT, NF, L,
            &      ACCEPT, AUX1, AUX3
            INTEGER NPORGRUP(1,GRUP),ASIGNA(1,N), SELEC(1,GRUP)
            REAL VAREST(IP, IP), SIMUL(N, IPK), VARCOV(IPK,IPK),
            &      B(IP,IP), B2(IP,IP), AUX2, CAMBIA, CHECA, CHECA2
            REAL, POINTER :: AUXI(:,)
            DATA ZERO /0.0/
C
            DO 1 I = 1, GRUP
1          NPORGRUP(1,I)= 0
C

```

```

AUX1=0
DO 2 I = 1, N
  IF (AUX1 .EQ. GRUP) THEN
    AUX1=0
  ENDIF
  AUX1 = 1+AUX1
  ASIGNA(1,I) =AUX1
  NPORGRUP(1,AUX1)= NPORGRUP(1,AUX1)+1
2  CONTINUE
C
CALL MATVARCOVA(IPK, N, SIMUL, VARCOV)
CALL FGPALG(VARCOV, B, IP, K, IPK, IFAULT, NF)
C
DO 3 I= 1, IP
  DO 3 J=1, IP
3   VAREST(I,J)=0
C
DO 4 I = 1, GRUP
  AUX1=0
  DO 18 M = 1, GRUP
    AUX2 = RAND()
    IF (AUX2 .LE. 0.5) THEN
      AUX1=AUX1+NPORGRUP(1,M)
      SELEC(1,M)= 1
    ELSE
      SELEC(1,M)= 0
    END IF
18  CONTINUE
  ALLOCATE (AUXI(AUX1, IPK))
  L =1
  DO 5 J = 1, AUX1
    ACCEPT = 0
    DO WHILE ((ACCEPT .EQ. 0) .AND. (L .LE. N))
      AUX3 = ASIGNA(1,L)
      AUX3 = SELEC(1,AUX3)
      IF (AUX3 .EQ. 1) THEN
        AUXI(J,:)=SIMUL(L,:)
        L=L+1
        ACCEPT =1
      ELSE
        L=L+1
      END IF
    END DO
5   CONTINUE
  CALL MATVARCOVA(IPK, AUX1, AUXI, VARCOV)
  CALL FGPALG(VARCOV, B2, IP, K, IPK, IFAULT, NF)
C
  DO 8 L= 1, IP
    CHECA2=0
    DO 8 J=L, IP
      CHECA=0
      DO 9 H=1, IP
        CHECA=CHECA+ABS(ABS(B2(H,J))-ABS(B(H,L)))
9     CONTINUE
      IF (CHECA2 .EQ. ZERO) THEN
        CHECA2=CHECA

```

```

                                ENDIF
                                IF (CHECA .LT. CHECA2) THEN
                                    CHECA2=CHECA
                                    DO 10 H=1,IP
                                        CAMBIA=B2(H,L)
                                        B2(H,L)=B2(H,J)
                                        B2(H,J)=CAMBIA
                                        IF (B2(H,L)*B(H,L) .LE. ZERO) THEN
                                            B2(H,L)=-B2(H,L)
                                        ENDIF
                                CONTINUE
10                                ENDIF
                                CONTINUE
8                                ENDIF
C                                CONTINUE
C
                                DO 6 L= 1, IP
                                    DO 6 J=1, IP
                                        VAREST(L,J)=VAREST(L,J)+(B2(L,J)-B(L,J))**2
6                                        CONTINUE
4                                CONTINUE
C
                                DO 7 I= 1, IP
                                    DO 7 J=1, IP
7                                        VAREST(I,J)=VAREST(I,J)/(GRUP)
                                END
C
C
C
C
C
C
C
                                CUERPO DE PROGRAMA
                                INTEGER IP, IPK, K, IFAULT, NF, GL, ENE, CORRECTO,
                                & GRUP1, GRUP2, GRUP3, SUP, DAT, SINO
                                REAL, POINTER :: A(:,,:), B(:,,:), FA(:,,:),
                                & F(:,,:), LAMBDA(:,,:),
                                & J1(:,,:), J2(:,,:), J3(:,,:),
                                & VB(:,,:),
                                & MU(:,,:), DATOS(:,,:),
                                & VAREST(:,,:)
                                CHARACTER(LEN=*) , PARAMETER :: FMT2 = "(500E16.4)"
                                CHARACTER(LEN=*) , PARAMETER :: FMT3 = "(A1,I4,A7,A2)"
                                REAL DETF, DETL, CHI, GAM, EME
                                CHARACTER*64 ARCHIVO
C
                                WRITE (*, *)
                                WRITE (*, *) 'ESTE PROGRAMA OBTIENE LOS ESTIMADORES DEL'
                                WRITE (*, *) 'MODELO DE COMPONENTES PRINCIPALES COMUNES'
                                WRITE (*, *) 'EN VECTORES DEPENDIENTES.'
                                WRITE (*, *)
                                WRITE (*, *) 'TECLEA INTRO PARA CONTINUAR'
                                READ(*,*)
                                WRITE (*, *)
                                WRITE (*, *) 'PARA ELLO SE REQUIERE TENER UN ARCHIVO DE TEXTO'
                                WRITE (*, *) 'DELIMITADO POR TABULACIONES QUE CONTENGA LA'
                                WRITE (*, *) 'BASE DE DATOS QUE SE QUIERE ANALIZAR, DONDE '

```

```

WRITE (*, *) 'CADA RENGLON TENGA LA SIGUIENTE ESTRUCTURA: '
WRITE (*, *) '(X11,X12,...,X1P,X21,...,X2P,.....XK1,...XKP)'
WRITE (*, *) 'SIENDO XIJ EL VALOR DE LA VARIABLE J EN EL GRUPO I.'
WRITE (*, *)
WRITE (*, *) 'TECLEA INTRO PARA CONTINUAR'
READ(*,*)
WRITE (*, *)
WRITE (*, *) 'ALTERNATIVAMENTE SE PUEDE TENER UN ARCHIVO DE TEXTO'
WRITE (*, *) 'DELIMITADO POR TABULACIONES QUE CONTENGA LA'
WRITE (*, *) 'MATRIZ DE VARIANZAS Y COVARIANZAS DE LOS DATOS,'
WRITE (*, *) 'TENIENDO LOS DATOS LA MISMA ESTRUCTURA ANTES'
WRITE (*, *) 'DESCRITA.'
WRITE (*, *)
WRITE (*, *) 'TECLEA INTRO PARA CONTINUAR'
READ(*,*)
WRITE (*, *)
WRITE (*, *) 'EL PROGRAMA OBTIENE LOS ESTIMADORES DE LAS MATRICES'
WRITE (*, *) 'BETA Y LAMBDA, LA ESTADISTICA JI-CUADRADA QUE '
WRITE (*, *) 'VALIDA EL MODELO CON SUS GRADOS DE LIBERTAD, LA '
WRITE (*, *) 'ESTIMACION DE LA MATRIZ DE VARIANZAS '
WRITE (*, *) 'ASINTOTICA DE BETA, ASI COMO LAS ESTIMACIONES DE '
WRITE (*, *) 'LA MATRIZ DE VARIANZAS DE BETA POR '
WRITE (*, *) 'LOS METODOS: GRUPOS ALEATORIOS, JACKKNIFE Y '
WRITE (*, *) 'REPLICACIONES REPETIDAS.'
WRITE (*, *)
WRITE (*, *) 'TECLEA INTRO PARA CONTINUAR'
READ(*,*)
WRITE (*, *)
WRITE (*, *) 'PARA GENERAR LAS ESTIMACIONES SE PUEDE OPTAR POR'
WRITE (*, *) 'EL SUPUESTO DE NORMALIDAD MULTIVARIADA O POR '
WRITE (*, *) 'EL SUPUESTO DE DISTRIBUCION ELIPTICA PEARSON'
WRITE (*, *) 'TIPO VII.'
WRITE (*, *)
WRITE (*, *) 'TECLEA INTRO PARA CONTINUAR'
READ(*,*)
WRITE (*, *)

```

C

```

WRITE(*,*) 'TECLEA 1 SI TU ARCHIVO CONTIENE LA BASE DE DATOS'
WRITE(*,*) 'TECLEA 2 SI TU ARCHIVO CONTIENE LA MATRIZ DE'
WRITE(*,*) 'VARIANZAS Y COVARIANZAS'
CORRECTO=0
DO WHILE (CORRECTO .EQ. 0)
    READ(*,*) DAT
    IF ((DAT .EQ. 1) .OR. (DAT .EQ. 2)) THEN
        CORRECTO = 1
    ENDIF
ENDDO
WRITE (*, *)
WRITE (*, *) 'DAME LA RUTA Y EL NOMBRE DE TU ARCHIVO.TXT'
WRITE (*, *) '(EJEMPLO: C: DCPC DATOS.TXT)'
READ (*, *) ARCHIVO
OPEN(UNIT=5,FILE=ARCHIVO, STATUS='old')

```

C

```

WRITE (*, *)
WRITE (*, *) 'DAME LA RUTA Y EL NOMBRE DEL ARCHIVO.TXT'
WRITE (*, *) 'DONDE QUIERES TUS RESULTADOS POR FAVOR: '
WRITE (*, *) '(EJEMPLO: C: DCPC RESULTADOS.TXT)'
READ (*, *) ARCHIVO
OPEN(UNIT=4,FILE=ARCHIVO)
C
WRITE (*, *)
WRITE (*, *) 'DAME EL NUMERO DE GRUPOS POR FAVOR: '
CORRECTO=0
DO WHILE (CORRECTO .EQ. 0)
    READ(*,*) K
    IF (K .GT. 0) THEN
        CORRECTO = 1
    ENDIF
ENDDO
C
WRITE (*, *)
WRITE (*, *) 'DAME EL NUMERO DE VARIABLES POR FAVOR: '
CORRECTO=0
DO WHILE (CORRECTO .EQ. 0)
    READ(*,*) IP
    IF (IP .GT. 0) THEN
        CORRECTO = 1
    ENDIF
ENDDO
C
WRITE (*, *)
WRITE (*, *) 'DAME EL NUMERO DE DATOS POR FAVOR: '
CORRECTO=0
DO WHILE (CORRECTO .EQ. 0)
    READ(*,*) ENE
    IF (ENE .GT. 0) THEN
        CORRECTO = 1
    ENDIF
ENDDO
C
IPK=IP*K
C
WRITE (*, *)
WRITE(*,*) 'TECLEA 1 SI TU SUPUESTO DISTRIBUCIONAL ES NORMAL'
WRITE(*,*) 'MULTIVARIADO'
WRITE(*,*) 'TECLEA 2 SI TU SUPUESTO DISTRIBUCIONAL ES PEARSON'
WRITE(*,*) 'TIPO VII'
CORRECTO=0
DO WHILE (CORRECTO .EQ. 0)
    READ(*,*) SUP
    IF ((SUP .EQ. 1) .OR. (SUP .EQ. 2)) THEN
        CORRECTO = 1
    ENDIF
ENDDO
C
IF (SUP .EQ. 2) THEN
    WRITE(*,*)
    WRITE(*,*) 'CUANTO VALE GAMMA?'
    CORRECTO=0

```

```

DO WHILE (CORRECTO .EQ. 0)
    READ(*,*) GAM
    IF (GAM .GE. 0) THEN
        CORRECTO = 1
    ENDIF
ENDDO
C
WRITE(*,*)
WRITE(*,*) 'CUANTO VALE M?'
CORRECTO=0
DO WHILE (CORRECTO .EQ. 0)
    READ(*,*) EME
    IF (EME .GE. 0) THEN
        CORRECTO = 1
    ENDIF
ENDDO
ENDIF
C
ALLOCATE (A(IPK, IPK),B(IP,IP),FA(IPK,IPK),
&    F(IPK,IPK), LAMBDA(IPK,IPK),
&    J1(IP*(IP-1)/2,IP*IP), J2(IP*(IP-1)/2,IP*IP),
&    J3(IP*(IP-1)/2,IP*IP),
&    VB(IP,IP),
&    MU(1, IPK), DATOS(ENE,IPK),
&    VAREST(IP,IP))
C
IF (DAT .EQ. 2) THEN
    WRITE(*,*)
    WRITE(*,*) 'TECLEA 1 SI CONOCES EL VECTOR DE MEDIAS DE'
    WRITE(*,*) 'LOS DATOS'
    WRITE(*,*) 'TECLEA 2 EN OTRO CASO'
    CORRECTO=0
    DO WHILE (CORRECTO .EQ. 0)
        READ(*,*) SINO
        IF ((SINO .EQ. 1) .OR. (SINO .EQ. 2)) THEN
            CORRECTO = 1
        ENDIF
    ENDDO
C
    IF (SINO .EQ. 1) THEN
        WRITE(*,*)
        WRITE(*,*) 'TECLEALO POR FAVOR (SEPARADO POR ESPACIOS)'
        READ(*,*) MU(1,:)
        WRITE(*,*)
    ELSE
        DO 8 I = 1,IPK
8           MU(1,I)=0
        ENDIF
    ENDIF
C
    WRITE(*,*)
    WRITE(*,*) 'EN EL METODOS DE REPLICACIONES REPETIDAS,'
    WRITE(*,*) 'CUANTOS SUBGRUPOS QUIERES USAR?'
    CORRECTO=0
    DO WHILE (CORRECTO .EQ. 0)
        READ(*,*) GRUP1

```

```

        IF (GRUP1 .GE. 0) THEN
            CORRECTO = 1
        ENDIF
    ENDDO
C
    WRITE(*,*)
    WRITE(*,*) 'EN EL METODOS DE GRUPOS ALEATORIOS,'
    WRITE(*,*) 'CUANTOS SUBGRUPOS QUIERES USAR?'
    CORRECTO=0
    DO WHILE (CORRECTO .EQ. 0)
        READ(*,*) GRUP2
        IF (GRUP2 .GE. 0) THEN
            CORRECTO = 1
        ENDIF
    ENDDO
C
    WRITE(*,*)
    WRITE(*,*) 'EN EL METODOS DE JACKKNIFE,'
    WRITE(*,*) 'CUANTOS SUBGRUPOS QUIERES USAR?'
    CORRECTO=0
    DO WHILE (CORRECTO .EQ. 0)
        READ(*,*) GRUP3
        IF (GRUP3 .GE. 0) THEN
            CORRECTO = 1
        ENDIF
    ENDDO
C
    IF (DAT .EQ. 1) THEN
        DO 9 J = 1, ENE
            READ(5,*) (DATOS(J, I), I = 1 , IPK)
            CALL MATVARCOVA(IPK, ENE, DATOS, A)
        ENDIF
C
    IF (DAT .EQ. 2) THEN
        DO 10 J = 1, IPK
            READ(5,*) (A(J, I), I = 1 , IPK)
        ENDIF
C
    CALL FGPALG(A, B, IP, K, IPK, IFAULT, NF)
C
    WRITE (4, *)
    WRITE(4, *)'LA MATRIZ BETA ES:'
    WRITE (4, *)
    DO 11 J = 1, IP
        WRITE(4, FMT2)(B(J, I), I = 1 , IP)
    WRITE (4, *)
    WRITE (4, *)
    WRITE(4, *)'EL NUMERO DE ITERACIONES UTILIZADAS ES:'
    WRITE (4, *)
    WRITE(4, *) NF
    WRITE (4, *)
C
    CALL KRONI(K, B, IP, IP, FA)
    CALL MULT2(A, FA, F, IPK, IPK)
    IF (SUP .EQ. 2) THEN
        F=((2*EME-IPK)/(2*EME-IPK-2))*F

```



```

ENDIF
CALL BLOQDIAG(F, IP, K, IPK, LAMBDA)
C
WRITE (4, *)
WRITE(4, *)'LA MATRIZ F ES:'
WRITE (4, *)
DO 13 J = 1, IPK
13   WRITE(4, FMT2)(F(J, I), I = 1 , IPK)
WRITE (4, *)
C
WRITE (4, *)
WRITE(4, *)'LA MATRIZ LAMBDA ES:'
WRITE (4, *)
DO 14 J = 1, IPK
14   WRITE(4, FMT2)(LAMBDA(J, I), I = 1 , IPK)
WRITE (4, *)
C
CALL DETERM(F, IPK, DETF)
CALL DETERM(LAMBDA, IPK, DETL)
C
CHI=ENE*LOG(DETL/DETF)
GL=(K*K-1)*IP*(IP-1)*0.5
C
WRITE (4, *)
WRITE(4, *)'LA ESTADISTICA JI CUADRADA ES:'
WRITE (4, *)
WRITE(4, FMT2) CHI
WRITE (4, *)
WRITE(4, *)'LOS GRADOS DE LIBERTAD SON:'
WRITE (4, *)
WRITE(4, *) GL
WRITE (4, *)
C
IF (IP .GT. 2) THEN
    CALL JACOB(B, IP, J1, J2, J3)
C
        CALL VARBNORM(J1, J2, J3, LAMBDA, B, IP, K ,IPK,
&                ENE, VB)
C
ENDIF
C
IF (IP .EQ. 2) THEN
    CALL JACOB2(B, IP, J1, J2)
C
        CALL VARBNORM2(J1, J2, LAMBDA, B, IP, K ,IPK,
&                ENE, VB)
C
ENDIF
C
WRITE (4, *)
WRITE(4, *)'LA MATRIZ DE VARIANZAS'
WRITE(4, *)'ASINTOTICA DE BETA ES:'
WRITE (4, *)
DO 15 J = 1, IP
15   WRITE(4, FMT2)(VB(J, I), I = 1 , IP)
WRITE (4, *)

```

```

C      IF ((SUP .EQ. 1) .AND. (DAT .EQ. 2)) THEN
          CALL SIMNORM(MU, A, IPK, ENE, DATOS)
      ENDIF
C      IF ((SUP .EQ. 2) .AND. (DAT .EQ. 2)) THEN
          CALL SIMPEAR(MU, A, GAM, EME, IPK, ENE, DATOS)
      ENDIF
      CALL REPLICA(DATOS, GRUP1, IPK, IP, K, ENE, VAREST)
      WRITE (4, *)
      WRITE (4, *) 'LA MATRIZ DE VARIANZAS DE BETA'
      WRITE (4, *) 'OBTENIDA POR EL METODO DE REPLICACIONES ES:'
      WRITE (4, FMT3) '(, GRUP1, ' GRUPOS', ' )'
      WRITE (4, *)
      DO 18 J = 1, IP
18      WRITE(4, FMT2)(VAREST(J, Z), Z = 1 , IP)
      WRITE (4, *)
C      CALL GRUPALEAT(DATOS, GRUP2, IPK, IP, K, ENE, VAREST)
      WRITE (4, *)
      WRITE (4, *) 'LA MATRIZ DE VARIANZAS DE BETA'
      WRITE (4, *) 'OBTENIDA POR EL METODO DE GRUPOS ALEATORIOS ES:'
      WRITE (4, FMT3) '(, GRUP2, ' GRUPOS', ' )'
      WRITE (4, *)
      DO 16 J = 1, IP
16      WRITE(4, FMT2)(VAREST(J, Z), Z = 1 , IP)
      WRITE (4, *)
C      CALL JACKKN(DATOS, GRUP3, IPK, IP, K, ENE, VAREST)
      WRITE (4, *)
      WRITE (4, *) 'LA MATRIZ DE VARIANZAS DE BETA'
      WRITE (4, *) 'OBTENIDA POR EL METODO DE JACKKNIFE ES:'
      WRITE (4, FMT3) '(, GRUP3, ' GRUPOS', ' )'
      WRITE (4, *)
      DO 17 J = 1, IP
17      WRITE(4, FMT2)(VAREST(J, Z), Z = 1 , IP)
C
C
C      WRITE (*, *)
      WRITE (*, *) 'TECLEA INTRO PARA FINALIZAR'
      READ(*,*)
C
      END

```